The Journal of

Applied Psychology

Vol. III JUNE, 1919 No. 2

EDITORS

G. STANLEY HALL JOHN WALLACE BAIRD

L. R. GEISSLER

And a Board of Co-operating Editors

WORCESTER, MASS.

FLORENCE CHANDLER, Publisher

Entered as second class matter June 9, 1917, at the post office at Worcester, Mass. under the Act of March 3, 1879.

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. V. BINGHAM, Carnegie Institute of Technology; JUNE E. DOWNEY, University of Wyoming; ROBERT H. GAULT, Northwestern University; WILLIAM HEALY, Baker Foundation, Juvenile Court, Boston; H. L. HOLLINGWORTH, Columbia University; JOSEPH JASTROW, University of Wisconsin; HERBERT S. LANGFELD, Harvard University; JAMES H. LEUBA, Bryn Mawr College; HENRY C. MCCOMAS, Princeton University; WALTER D. SCOTT, Northwestern University; C. E. SEASHORE, University of Iowa; DANIEL STARCH, University of Wisconsin; EDWARD K. STRONG, JR., George Peabody College for Teachers; L. M. TERMAN, Leland Stanford Jr. University; E. L. THORNDIKE, Columbia University; J. E. WALLACE WALLIN, Psycho-Educational Clinic, St. Louis; G. M. WHIPPLE, Carnegie Institute for Technology, Pittsburgh; HELEN THOMPSON WOOLLEY, Vocation Bureau, Cincinnati.

The scope of the JOURNAL OF APPLIED PSYCHOLOGY will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis (c) The influence of general environmental conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, etc.

VOLUME III JUNE, 1919 NUMBER 2

TABLE OF CONTENTS

V. A. C. HENMON. Air Service Tests of Aptitude for Flying 103

L. L. THURSTONE. Mental Tests for Prospective Telegraphers; A Study of the Diagnostic Value of Mental Tests for Predicting Ability to Learn Telegraphy 110

GRACE ARTHUR and HERBERT WOODROW. An Absolute Intelligence Scale: A Study in Method 118

S. L. PRESSEY and L. W. PRESSEY. "Cross-out" Tests, With Suggestions as to a Group Scale of the Emotions 138

GEORGE W. FRASIER. A Comparative Study of the Variability of Boys and Girls 151

WALTER B. SWIFT. Mental Attitudes in Debate 156

JOSEPHINE CURTIS FOSTER. A Case of Intellectual Development Despite Enforced Seclusion 167

F. L. WELLS and C. M. KELLEY. Briefer Studies from the Psychological Laboratory of McLean Hospital 172

NOTES AND NEWS 194

Communications and manuscripts may be sent to any one of the three editors; reviews, summaries, notices, books and off-prints for review should be addressed to Dr. L. R. Geissler, Clark University, Worcester, Mass. Contributors are requested to follow the conventional method of citation employed by this JOURNAL; in every case the year, volume and pages should be specified.

THE JOURNAL OF APPLIED PSYCHOLOGY is published quarterly. The subscription price is $4.00 a year. Single numbers $1.25. Remittances and business communications should be addressed to Florence Chandler, Publisher, Clark University, (950 Main St.), Worcester, Mass.

THE
JOURNAL OF APPLIED PSYCHOLOGY

VOL. III JUNE, 1919 No. 2

AIR SERVICE TESTS OF APTITUDE FOR FLYING[1]

By V. A. C. HENMON, University of Wisconsin

From the psychologist's point of view an outstanding feature of the war has been the demand for highly specialized abilities and aptitudes in addition to general military virtues. No where has this been more true than in the Air Service and no where have the problems of selecting individuals and assigning them to the work for which they are best fitted been more urgent. As late as April 1917, the Aviation Section of the Signal Corps had but 52 trained flyers. The Air Service with upwards of 16,000 flyers is thus a new product.

When war was declared the need for selecting thousands of men for training as pilots, observers and balloonists was met by the creation of Examining Boards and Physical Examining Units. Rigorous physical and medical standards were adopted. While the mental, moral and professional requirements were equally exacting the standards were necessarily vague and general. We were instructed to select men of good education and high character, men who were in every way qualified and fitted to become officers of the U. S. Army — a rather intangible set of specifications. We were constantly enjoined to remember that the flying officer was not to be an "aerial chauffeur," but a "twentieth century cavalry officer mounted on Pegasus."

Examiners soon found that while it was relatively easy to pick men who could negotiate the non-Pegasarian flights demanded in the Ground Schools, what they would do when

[1] Paper prepared for joint meeting of Section H of the A. A. A. S. and American Psychological Association at Baltimore. Delay in securing approval of War Department prevented the reading of the paper.

they mounted Pegasus was a sheer guess. The result was that in spite of the Examining Boards and Physical Examining Units which eliminated 50 to 60 per cent of applicants, and in spite of Ground Schools, which eliminated 15 per cent of the survivors, there were still at least 6 per cent of men who reached Flying Fields who were discharged or transferred because of inaptitude for flying. This estimate of 6 per cent was obtained from independent data obtained by Thorndike and myself. Thorndike's estimate is based on a follow-up of 1,600 cadets who were admitted to Flying Schools and were either commissioned or discharged from flying status or killed in accident. Observations at Kelly Field showed that on June first, when 750 Reserve Military Aviators had been commissioned, 74 men had been discharged. Of these 55 were discharged for inaptitude or temperamental unfitness for flying, 11 for physical disability or incapacity, 5 for misconduct and 3 for technical reasons. In at least half the cases of discharge for physical unfitness this is but a euphemism for temperamental unfitness for flying. There were about 600 cadets at the Field at the time. Estimating the discharges or transfers that would accrue from this number the estimate of 6 per cent is conservative for Kelly Field. The recent report of the Director of Military Aeronautics for the period ending June 30th, showed that at that time 4,980 men had been commissioned and about 400 had been dropped for inability to fly.

Over and above the 6 per cent who failed completely there are a considerable number who, while ultimately commissioned, have no aptitude for flying, learn with difficulty and smash ships. For example, there were two men who were rated as very poor in flying and who did very poorly in the tests. One took 106 hours for his R. M. A. when the average time was well below 50 hours, and the other took 94 hours. These men were commissioned because it seemed better to use them than to discharge them and train others at great cost in their stead. The 74 men discharged wasted 1,400 hours of flying. Some statistician has figured the total cost of flying training as $2.00 per minute in the air. Whether this includes the cost of wrecked ships is not known, but this group wrecked not less than 30 ships. One man, for example, was given 55 hours of flying instruction and wrecked five ships before he was relieved as "unadaptable for further training." The money cost of the trial and error method of discovering ability to fly, to speak merely of it, becomes enormous. It is unnecessary to multiply evidence to show that the need for improvement in methods for selecting flyers was and is very great.

In June 1917 the National Research Council Committee on Aviation tried out tests reported as promising by French and Italian psychologists, and other tests as well, on 75 candidates at the Massachusetts Institute of Technology Ground School. This work was done by Burtt, Troland and Miles. Stratton, in August 1917, tried out tests on 50 flying cadets at San Diego, and Maxfield tested 44 cadets at Essington Field. Dr. Parsons in the early part of this year tested the cadets in the Naval Aviation Service at Massachusetts Institute of Technology. In this way some forty odd tests had been given something of a try-out. The difficulty of securing satisfactory ratings of flying ability or of an adequate follow-up made the correlations, however, too uncertain for practical conclusions.

In March of this year Captain Stratton and I were detached from Examining Boards to make, in coöperation with Thorndike of the Committee of Classification of Personnel, an official investigation of the most promising of the tests as shown in earlier studies and of others which seemed *a priori* to be worth investigation. It would be too long a story to list the forty tests, give the correlation with flying ability which they showed, and indicate the grounds of selection of tests for further try-out. Ability to fly would seem to require emotional stability and control or "nerve," sensitiveness to tilt, general motor coördination and control, quickness of response to visual and auditory stimuli and to changes of equilibrium, mental alertness, and athletic ability and interest. Tests of most of these traits had shown promise. These, and several others that gave no results worth noting here, were therefore selected and given to 300 cadets and flying instructors at Rockwell Field by Stratton and by myself at Kelly Field. At each field 50 very good flyers, 50 very poor flyers, some of whom had already been relieved from further training for inability to fly, and 50 unknowns were given the tests. Each man was tested on two different days. The tests selected and given at both fields were as follows:

1. *Emotional Stability,* a test which measures the liability of the subject to incoördinate purposeless reaction upon a receipt of a sudden shock, the discharge of a pistol. Measurements were made of amplitude of hand tremor, time of recovery from tremor, changes of frequency of pulse, changes in frequency and amplitude of respiration, and the effect of the shock on rate of solving examples in addition. Of these measurements amplitude of hand response and changes in frequency in respiration were found to be especially significant.

2. *Perception of Tilt,* which measures by means of a

specially constructed tilting chair, the sensitiveness to *gradual* changes of bodily position.

3. *Swaying,* a test of the motor coördination and control required for the subject to stand steadily with the eyes open and with the eyes closed.

4. *Visual Reaction,* quickness of response to movement of the upper hand on the dial of the Hipp chronoscope.

5. *Auditory Reaction,* quickness of response to the noise of the Zimmerman sound hammer.

6. *Equilibrium Reaction,* quickness of response to *sudden* changes of bodily position, to the right and to the left. Burtt's tilting table, which gives a negatively accelerated tilt of constant distance and speed, was used.

7. *Equilibrium Differential,* the sum of the visual and auditory reaction times minus the equilibrium reaction time, a measure of the relative quickness of the response to changes of equilibrium as compared with simple reaction times.

8. *Extension of Curves,* a test roughly simulating the judgments of distance required in landing a ship. This test which showed promise in the earlier study by Stratton gave negative results in the later investigation.

In addition to the psycho-physical tests two tests by means of blank forms were given.

9. *Thorndike's Mental Alertness Test,* an arrangement of thirteen tests of various functions designed to measure native mental ability or general intelligence. This test had been adopted for use by Examining Boards and formed a regular part of examination procedure.

10. *Athletic Achievement and Interest.* Thorndike's study of application blanks filled out by candidates and subsequent performance had shown a positive correlation between athletic ability and success in flying. A more detailed blank was prepared and a scoring system for it worked out by Thorndike and Kelly.

In view of the fact that the Mental Alertness and Athletic Achievement tests were not given to enough men at Rockwell Field, and since they are in the series of tests recommended and adopted, I shall discuss the results from Kelly Field mainly. The instructors and cadets were rated in flying ability by the Officers in Charge of Flying on a basis of judgments of five stage commanders and the progress cards. All officers showed keen interest in the problem and the ratings were made with great care.

The correlations with flying ability, the inter-correlations of each test with every other test, and the intra-correlations

between the trials on two days were determined. Taking into account all available data as well as the later investigations the correlations of the individual tests with flying ability are as follows:

CORRELATIONS WITH FLYING ABILITY

		K. F.
1. Emotional Stability,		
1. Hand Response	.35	.26
2. Changes in Freq. of Respiration	.25	.23
2. Perception of Tilt	.23	.26
3. Swaying	.22	.16
(4. Visual Reaction	0	.15)
5. Auditory Reaction	.14	.15
6. Equilibrium Choice Reaction	.15	.08
7. Equilibrium Differential	.18	—.15
(8. Extension of Curves	—.03	.14)
9. Mental Alertness	.35	.23
10. Athletic Achievement		

The correlations in the first column were obtained by Shepard's formula. Those in the second column are the correlations from the Kelly Field data by the rank method. The figures in parenthesis are for tests not in the recommended series.

The best tests, as the table shows, are emotional stability, especially the hand response, perception of tilt, and mental alertness. "With the material used the correlation between the Thorndike-Kelly athletic score in flying ability was over .6 though certain factors were operative which make this possibly too high and the number of cases was under 100." None of these tests alone obviously is a sure symptom for inaptitude for flying or a sufficient basis for prognosis. When, however, the individual measurements are combined the composite score obtained from the series or team of tests recommended gives a multiple correlation coefficient of .70, which is a sufficient basis for prognosis with some confidence. In selecting the team of tests those were chosen which show the highest correlations with flying ability, satisfactory reliability coefficients, and low correlations with each other. To determine the weights to be given each test toward the composite score would ideally require the calculation of partial correlations and coefficients of regression. The time and labor for this analysis has not been available. A provisional scoring and weighting scheme, scores being expressed as deviations above and below the mean of performance in the tests, was worked out. The weights were estimates made with due regard to the principles and factors that determine weighting by the partial correlation method. The success with which the recommended

tests segregate the poors and the goods is indicated in the following table:

SCORES ON FLYING ABILITY OF 100 CADETS AND INSTRUCTORS AT KELLY FIELD

Score in the recommended series of tests	Report from flying school p-poor. g-good.	Notes to date of June 15	Average dual time of others
—75 or worse	pppppp	4 suspended, relieved or discharged	15:50
—50 to —74	ppppppg	4 " " " "	11:16
—25 to —49	pppppppp	5 " " " "	10:17
	ppppppg		
0 to —25	pppppggg	9 " " " "	10:03
	ppppggg		
	ppppggg		
	pppp		
0 to —25	ppggggg	1 " " " "	9:19
—25 to —49	pggggggg	1 not reported	9:23
	pggggggg		
—50 to —74	ggggggg		7:50
—75 or better	g		7:21

The acid test is, of course, the ability to predict. In order to eliminate the possibility of confusing abilities produced by flying with those prophetic of ability to fly, the tests were given to 50 cadets of unknown ability who had just reached the Field. Some of these had had no experience in flying and none over two or three hours of dual instruction. On a basis of performance in the tests solely I submitted to the Officer in Charge of Flying a prediction that five of these men would either be discharged or learn to fly with great difficulty. Two others were selected as showing special aptitude, one of whom because of deficiency in formal education had difficulty in getting into the Service. Of the five men three were discharged after 4, 20 and 22 hours of instruction respectively. One of these men wrecked two ships completely. The fourth man was relieved from flying at one time and brought before the Discharge Board but was given another chance and finally was commissioned after 85 hours of flying instruction when the median time for his class was 60 hours. The fifth man was commissioned after 93 hours of instruction. The Officer in Charge of Flying said of these men that they did "Fair" work. The two men who were picked to show exceptional aptitude were commissioned as pursuit pilots and their work characterized as "Very Good." One of the men made remarkable progress and was commissioned after 43 hours of flying when the median time was 70 hours. This was the man who barely met the educational standards.

The success with which the tests differentiated the goods from the poors and the correspondence between prediction and performance were sufficiently convincing so that the Personnel Section of the Air Service was authorized to proceed with their introduction. Steps were taken to organize four Personnel Units to be established at the four Ground Schools and in affiliation with four Examining Boards. The all too familiar vexatious delays in securing necessary personnel and equipment prevented their being fully organized and actually in operation before the armistice was signed and Examining Boards and Ground Schools were closed.

While waiting for personnel and equipment at Princeton, where the training unit was located, Captain F. L. Wells and I carried out a series of experiments with four subjects to determine (1) the susceptibility of the tests to systematic practice, (2) the advisability of repeating the tests on the same day instead of as heretofore on the next or some subsequent day, and (3) the effect of time of day on performance in the tests. In the practical administration of the tests questions on these points would naturally arise and did arise. The tests, except those of mental alertness and athletic achievement, measure relatively simple functions that enter into flying ability. *A priori* one might expect, though one could not be sure without actual experiment, that these tests, having been highly practiced by the ordinary conditions of life, would not show very quick or very marked effects of practice or of diurnal variation. If this supposition holds, the tests are for practical purposes all the more valuable. The results show

1. That except for the equilibrium choice reaction practice effects are slight for daily tests over a period of fifteen days. The practice effects, moreover, are least evident in those with the poorest initial scores.

2. That the variability in the once-a-day tests is but little greater than in the twice-a-day tests.

3. That time of day for all practicable purposes is negligible.

The Air Service is certain to become an increasingly significant arm of the military forces and it is equally certain that aviation will have ere long important applications in civil life. The selection of the apt and the elimination of the unfit for flying is, therefore, so important that the responsible authorities should on a basis of the showing made be required to give the tests a practical tryout and take steps for their improvement, in order that vast sums should not continue to be expended fruitlessly and human life wasted.

MENTAL TESTS FOR PROSPECTIVE TELEGRAPHERS, A STUDY OF THE DIAGNOSTIC VALUE OF MENTAL TESTS FOR PREDICTING ABILITY TO LEARN TELEGRAPHY

By L. L. THURSTONE, Carnegie Institute of Technology, Pittsburgh

This investigation was made to ascertain whether it is possible to predict ability in telegraphy by mental tests. Our subjects were 165 drafted men of Class I, who registered for a night course in radio-telegraphy. The course was given at the Carnegie Institute of Technology to prepare drafted men for the Signal Corps of the U. S. Army. The course opened on November 19, 1917, and this report is written in March, 1918.

The two main features in this investigation are the mental tests and the daily records of progress. The mental tests were given to the men immediately after registration for the course on the same evening that they were given the physical examination. Every applicant took both the mental and physical examination before entering classes. The records of progress were obtained by means of prescribed tests of receiving speed. The instructions and copy for the tests of speed were prepared in the office of the Psychology Department and administered to all of the fourteen classes with as much uniformity as possible. The receiving speed test was of two minutes duration, given at every meeting of the class and always between the first and second hours of the three-hour evening session. A more complete description of these tests and the directions for administering them are contained in my "*Handbook for Teachers of the Telegraphic Code,*" which has been published as Bulletin No. 16 of the Federal Board for Vocational Education, Washington, D. C.

I. SUBJECTS

1. *Occupations.* Among the 165 students were represented the mechanical and electrical trades, building and steel trades, engineers, chemists, salesmen, bookkeepers, etc. It is of interest to note that there is no occupational differentiation in the ability to learn telegraphy. Hence, occupation is not suitable as an index for selecting prospective telegraphers.

2. *Schooling.* Here, again, we find the rather surprising fact that schooling is not suitable as an index of ability to learn telegraphy as far as this type of group is concerned. It is quite likely that the factor of schooling is loaded in this group of students so that the brightest of the high school and college graduates may have found preferred service in the army before being drafted. If this is true, then the college, high school, and grammar school groups here represented are not comparable. At any rate, if we take our findings at face value, ability in telegraphy cannot be predicted on the basis of general schooling.

The Pearson correlation coefficient between number of years of schooling and receiving speed in telegraphy after one hundred hours of practice is 0.00 ± 0.08; ($n=158$). While the correlation coefficient is low, it may be significant that the eleven students who left school before finishing the eight grades were conspicuously low in telegraphy.

3. *Age.* Being a group of drafted men the age limits were 21 to 31 years. The frequencies for the higher ages within the draft limits are relatively low. This is explained by the greater frequency of married men at those ages and by other causes of exemption among the older of the men of draft age. It is not without interest to note that the older men exhibited more perseverance in staying by the rather monotonous evening buzzer practice than the younger men.

The correlation between age and receiving speed in telegraphy after three months practice (about 100 hours) is -0.09 ± 0.08, which indicates that age has no diagnostic value in predicting ability in telegraphy within the age limits of 21 and 31.

II. THE MENTAL TESTS

All of the students who entered the radio course during November and December took a psychological examination consisting of eight mental tests. The purpose of these tests was to ascertain whether mental tests could be used to predict ability to learn telegraphy.

We gave the following tests:

1) Rhythm Test
2) Opposites Test
3) Analogies Test
4) Gordon Directions Test
5) Trabue Completion Test
6) Spelling Test
7) Arithmetic Test
8) Sentence Test

All of these mental tests have been checked up with the highest receiving speed attained during the first one hundred hours of practice.

1. *Rhythm Test.* This test was designed specially for the purpose of testing ability in telegraphy. It is given as follows: The examiner is equipped with a buzzer and telegraphic key. The examiner sounds on the buzzer a rhythm pattern consisting of auditory dots and dashes. Each pattern is repeated once in order to give all students a fair chance to hear it. They then reproduce the pattern, using a short horizontal line for a dash and a short vertical line for a dot. The test consists of thirty-five of these rhythm patterns given in the order of difficulty, beginning with the easiest. The score in the rhythm test is the number of errors made. The rate of presentation is such as would correspond to a telegraphic sending speed of ten words per minute and the whole test is given in about fifteen minutes.

The rhythm test gives a correlation of $+0.48 \pm 0.06$. Of the highest thirteen students in this test, only one did worse than the average in telegraphy. *Of the lowest fifteen students in this test, only three students attained the average of the class.*

2. *Opposites Test.* This is a good test for predicting ability in telegraphy. The correlation coefficient for speed in receiving telegraphic code and scores in the Opposites Test is $+.42 \pm 0.04$.

The ten radio students who rank lowest in the Opposites Test are all inferior in telegraphy. Not one of them attained even average performance in telegraphy. Their average receiving speed is 3.8 words per minute as compared with 6.4 words per minute for the whole group.

The lowest twenty per cent of the group (23 students), arranged according to the Opposites Test, have an average receiving speed of 4.5 words per minute as compared with 6.4 for the whole group. Of these 23 students, only five exceeded the average of the group in telegraphy.

3. *Analogies Test.* The Pearson correlation coefficient for receiving speed in telegraphy and the Analogies Test is $+0.29 \pm 0.05$.

Analyzing the scatter diagram more closely, we find that *of the seventeen students who constitute the lowest ten per cent of the group, according to the Analogies Test, only one student attained the average of the class in telegraphy.* The average speed in telegraphy for these seventeen students is 5.2 after one hundred hours of practice. The average speed in teleg-

raphy for the upper ten per cent of the group, divided according to the Analogies Test, is 7.5 words per minute after one hundred hours of practice. The average speed in telegraphy for the whole group at the same stage of practice was 6.4 words per minute.

This indicates clearly that while the Analogies Test does not give a high correlation coefficient it can be used to eliminate a considerable portion of those who lack the special ability to learn telegraphy. The average telegraphic speed for the lowest half of the group, divided according to the Analogies Test, is 5.4 words per minute, the speed for the upper half of the group is 7.1 words per minute.

If the lowest twenty per cent of the group had been eliminated on the basis of the Analogies Test the average speed in telegraphy would have been 7.0 words per minute instead of 6.4 words per minute.

4. *Gordon Directions Test.* The correlation of this test with speed in telegraphy after one hundred hours of practice is $+0.27 \pm 0.05$. It has diagnostic value for ability in telegraphy for the lowest ten per cent of the group arranged according to the test, but not for the middle and higher scores in the test. *Of the 33 radio students who scored in the lowest twenty per cent of the group, only four succeeded in attaining the average of the class in telegraphy.* The average receiving speed for this lowest twenty per cent of the group, divided according to the test, was 4.6 words per minute, the average for the whole group was 6.4 words per minute. If those registrants who scored in the lowest twenty per cent of the group had been eliminated at the start the average receiving speed for the class would have been raised from 6.4 words per minute to 6.9 words per minute.

If those radio students who were in the lowest half of the group according to the Directions Test had been eliminated the average receiving speed of the class after 100 hours of instructions would have been raised from 6.4 words per minute to 8.1 words per minute.

The two most important facts about the use of the Directions Test for selecting telegraphers are as follows:

(1) *If we consider all those students as failures who did not exceed three words per minute in receiving tests after 100 hours of practice, the test would enable us to exclude 48 per cent of these failures;* (2) only 12 per cent of those who score below 40 in the test succeed in doing better than the average of the class.

5. *Trabue Completion Test.* We used the Bureau of Salesmanship form of this test. It gives a correlation with telegraphy of $+0.21 \pm 0.05$. Of the thirteen lowest students in this test, four students attained the average of the class in telegraphy. Of the eleven highest students in this test, only two were below the average in telegraphy. While this test is not without some diagnostic value for ability in telegraphy its best use is as a supplement to other more successful tests. This test is usually a good index of general intelligence, but in the present investigation our criterion is ability in telegraphy which is a special ability and that probably accounts for the rather low rank of the completion test.

6. *Spelling Test.* This test was given to the radio students in the belief that ability in spelling might prove to be of diagnostic value in the selection of telegraphers. The correlation between speed in telegraphy and score in the spelling test is $+0.18 \pm 0.06$.

Although this coefficient is lower than one would expect, the spelling test can be used effectively for eliminating some failures in telegraphy. *Of the 12 students who made the lowest scores in spelling, only two attained the average of the class in telegraphy.* The average speed in telegraphy for these twelve students was 4.8 words per minute as compared with 6.4 words per minute for the whole group.

7. *Arithmetic Test.* The correlation of this test with speed in telegraphy is only $+0.08 \pm 0.05$. Inspection of the scatter diagram shows that this test has no diagnostic value for ability in telegraphy.

8. *Sentence Test.* This test gives a correlation with telegraphy of $+0.09 \pm 0.05$, which indicates that the test has no diagnostic value for ability in telegraphy.

III. SUMMARY

(1) Figure 1 shows in black the speed in telegraphy of those who failed in any two or more of seven tests. The seven tests here considered are: Rhythm, Opposites, Analogies, Directions, Sentence Completion, Spelling, Number Completion. The white spaces indicate the telegraph speed of those who did not fail in more than one test.

(2) Figure 2 shows the scatter diagram for the Rhythm Test with a Pearson correlation coefficient of $+0.48 \pm 0.06$.

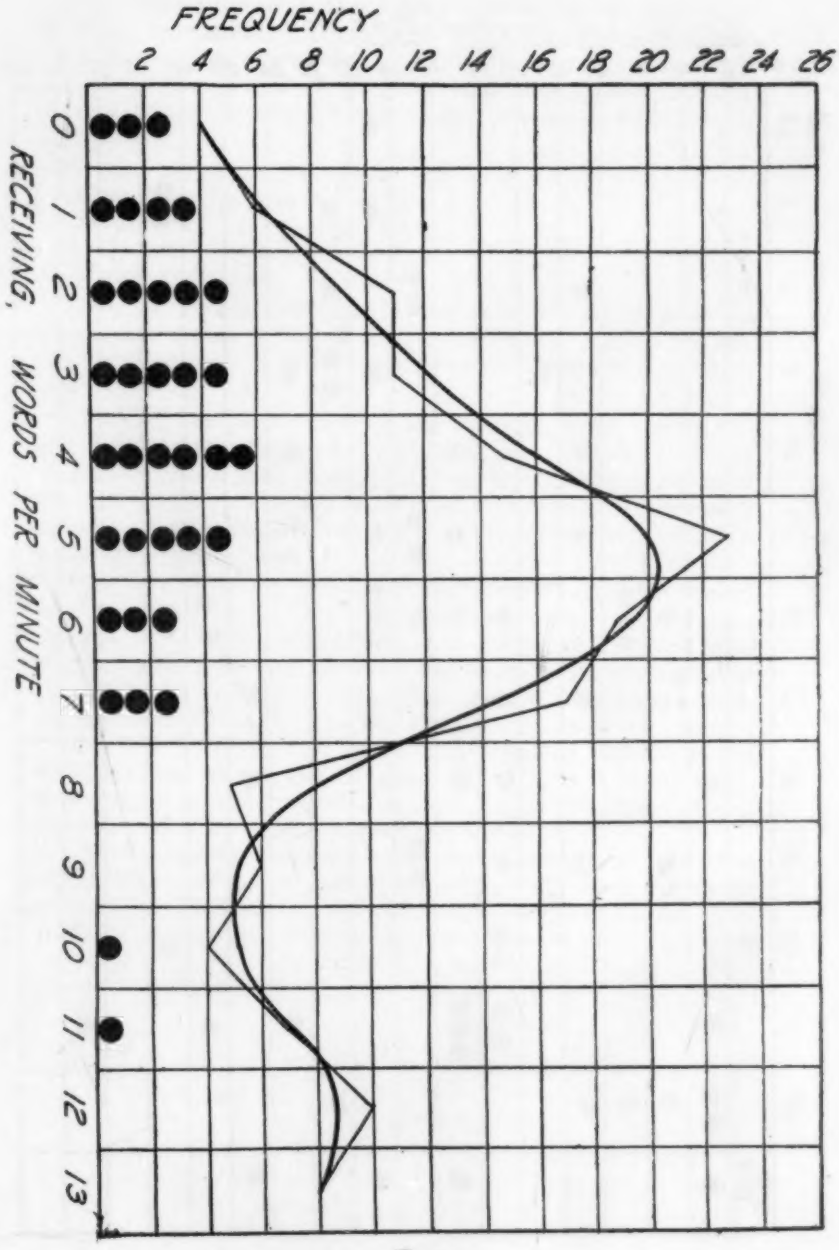

FREQUENCY
2 4 6 8 10 12 14 16 18 20 22 24 26
RECEIVING, WORDS PER MINUTE
0 1 2 3 4 5 6 7 8 9 10 11 12 13

FIG. 1

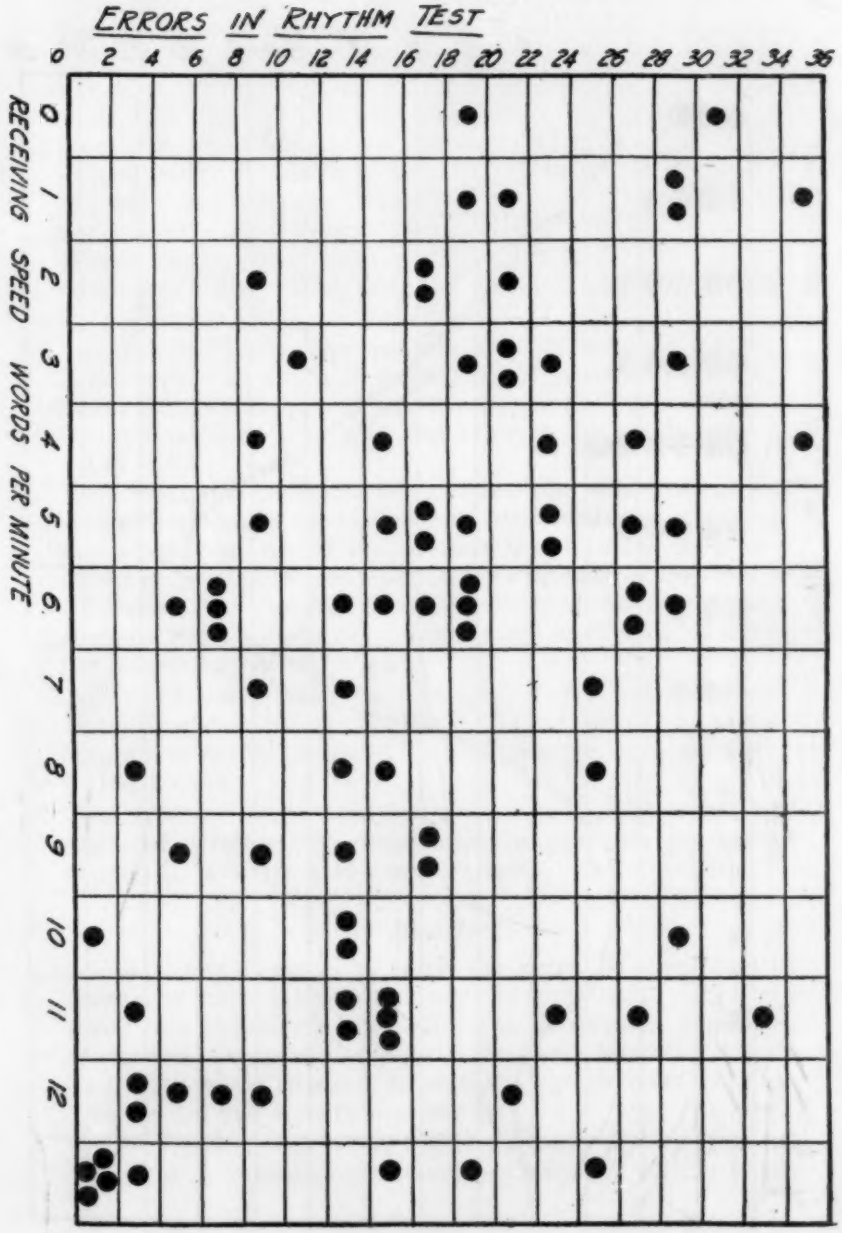

ERRORS IN RHYTHM TEST
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
RECEIVING SPEED WORDS PER MINUTE
0 1 2 3 4 5 6 7 8 9 10 11 12

FIG. 2

(3) Five of the tests have been evaluated by the method of multiple correlation: These tests were: Rhythm, Opposites, Completion, Analogies, and Gordon Directions. The fourth order coefficients for these tests with telegraphic receiving speed after one hundred hours of practice as the criterion are as follows:

Rhythm: $r_{12\,(3456)} = +.31$
Opposites: $r_{13\,(2456)} = +.21$
Completion: $r_{14\,(2356)} = +.096$
Analogies: $r_{15\,(2346)} = +.02$
Gordon Directions: $r_{16\,(2345)} = +.06$

(4) The general intelligence tests are not as valuable for diagnosing ability to learn telegraphy as for measuring general intelligence. Ability in telegraphy is probably a special ability.

(5) The fact that years of schooling does not agree with ability to learn telegraphy indicates that this is a special ability. College graduates usually do better on general intelligence tests than those who have only finished grammar school. But college graduates do not necessarily excel in learning telegraphy.

(6) The total correlation coefficient $R_{1(23456)}$ for receiving speed in telegraphy on the one hand and the combined score in five tests is $+.53 \pm 0.04$. The Rhythm Test alone gives a zero order coefficient of .48 which is not noticeably raised by the addition of four of our best general intelligence tests. Hence the Rhythm Test carries most of the load.

(7) The bimodality of figure 1 is probably significant and indicates that we are here dealing with a special ability. The upper mode probably represents that part of the group of subjects in which the special ability is present.

(8) A vocational survey of students in a radio mechanics course shows that schooling is one of the best diagnostic criteria in selecting men to be trained in the care and repair of wireless apparatus but our analysis also shows that when selecting men to be trained as telegraph operators the schooling is of no diagnostic value.

AN ABSOLUTE INTELLIGENCE SCALE: A STUDY IN METHOD

By Grace Arthur and Herbert Woodrow, University of Minnesota

Aim: The present investigation[1] is a study of a series of group tests as regards their value,—

(1) in classifying children of grade school age according to mental ability;

(2) in the construction of a scale of absolute intelligence units;

(3) in comparing the "more or less" with the "all or none" method of scoring; and

(4) in a system of weighting, which is neither arbitrary nor determined by subjective evaluation of the tests used, but which is logical in its origin and automatic in its application.

The aims of this investigation need but little explanation. The need for group intelligence tests has long been recognized and several systems of such have made their appearance since the present investigation was undertaken. Our interest, however, is not primarily in the development of group tests, but in certain fundamental features of method, of the same importance in individual testing as in group testing. The feature in which we are mainly interested is the use of absolute intelligence units,—units, that is, which are equal at all points of the scale and in terms of which it is possible to plot an absolute intelligence curve showing (in so far as the tests used really measure intelligence) the absolute increase in intelligence of children with increase in chronological age.

In the selection of tests, we have chosen only such as permit a measurement of degree in the ability tested, whatever that degree of ability may be. Such tests offer a marked contrast with the Binet-Simon tests. Many of the latter determine merely whether or not a certain degree of ability is exceeded, instead of measuring by how much it is exceeded or by how

[1] It is with pleasure that acknowledgment is made to Dr. Frances Lowell, of the University of Minnesota for suggestions as to the choice of tests and details of procedure, and to Miss Edith Taylor, principal of the Hancock School (the school tested), St. Paul, Minn., for her aid in securing a favorable attitude towards the tests on the part of teachers and pupils.

much it fails of being reached. Such tests, the child either fails or passes. They may be termed "all or none" tests. In contradistinction to them, the tests here employed may be called "more or less" tests, because they measure the ability which they test in any amount from zero up to the maximum.

When scored by the more or less method, different tests, such as must be used in any measurement of intelligence, give measures in different units. It becomes an important problem, consequently, how to average together the results of the different tests. A common method of combining the results of several tests is to weight them, but the methods of weighting heretofore used appear unsatisfactory,—and even, in some conspicuous cases, entirely arbitrary. We have found, however, that by using absolute intelligence units of the sort mentioned, weighting of the different tests occurs logically and automatically, and in exact correspondence to their value in discriminating between children of different age and, therefore, presumably of different intelligence.[2]

Choice of Tests: As the tests were to be given to all the pupils of a grade school, it was necessary to select such as would be simple enough for the younger children to comprehend, and at the same time difficult enough to tax the ability of the most mature. The tests finally decided upon were the following:

Immediate Memory Span (digits)
Easy Opposites (new lists based on the data obtained by King and Gold)[3]
Hard Opposites
Substitution (Woodworth—Wells)
Word Building (*a, e, i, r, l, p,* and *a, e, o, b, m, t*)
Language Completion (Trabue, Scales B and L)
Anagrams (original)
Cancellation (three letter)
Comprehension (adapted from Kuhlmann)

The only ones of these tests requiring special description are the anagrams test and the comprehension test. The anagrams test consisted of twenty-one words, varying in length from three to seven letters, taken from the lower end of the Ayres' spelling scale. A definite scheme of letter displace-

[2] It hardly needs to be pointed out here that the only natural intelligence scale by which the validity of psychological scales can be checked is that formed by the chronological age groups.

[3] The Standardization of Certain Opposites Tests. *Journal of Educational Psychology,* 1916, p. 459.

ment was followed and the words were arranged in order of difficulty estimated from their relative length and the number of letters displaced. This trial list was given to the pupils of the first, second, sixth and eighth grades of another school, and the words were then arranged in increasing order of difficulty as shown by a frequency table based upon the responses of these pupils. Spelling ability, as a factor of success in this test, was reduced to a minimum by using only the easiest words from the Ayres' list. To offset the puzzle element as much as possible, the subjects were instructed, in case they came to a word they could not get, to go on to the next and to come back to the hard word later.

For the comprehension test, three of the best individual tests from the Kuhlmann 1917 revision[4] were combined and adapted for group work. Two test sheets were used, the easier for all grades, the harder for the 4th to the 8th grades inclusive. Where results from both forms were used, they were added to give the total score.

Subjects: The subjects were grade school pupils from 6 to 17 years of age, inclusive. Norms for the separate tests were secured for the ages 6 to 13 years (though the system of tests here employed is not regarded as adapted to children under 7 years). Results for the age groups above 13 were discarded, because all the members of these groups were retarded pedagogically. Eleven subjects of the 13 year group were one-half to one year advanced pedagogically, and as they entered high school before the testing was completed, they had to be rejected from the cases finally used. From an examination of the age and grade distribution of the school, it would appear that the 10 year group had fewer bright subjects than had the other age groups. It is probable, therefore, that the 10 years norms, as well as the 13 year norms, are too low. This is of little concern to us in the present stage of our investigation,[5] as we believe we can improve on the selection of tests used and expect to obtain new norms.

For all subjects, chronological ages were reckoned to March 8th, 1918, as that date would give an error of less than one month in the great majority of the cases.

Instructions: Written instructions were used by the experimenter for all the tests, and were half spoken, half read. For the most part, the instructions were adaptations of those of

[4] Kuhlmann, The Measurement of Mental Development. *Journ. of Psycho-Asthenics, Monog. Sup.*, 1917.

[5] This work was interrupted by the entrance of one of the writers into military service, but will be continued by the other.

earlier investigators. They were given a preliminary trial with first, second, sixth and eighth grade groups of another school, and were then revised on the basis of the results with these groups.

In giving instructions, group response in a trial test was secured wherever possible. Purity of diction was sacrificed in every case where a colloquial form appeared to be more readily understood. We spare the reader details of the tests and instructions, hoping to present them at a future time when our scale of mental measurement is more perfectly standardized.

Norms: With each age group, the mean, as the most probable true value, was taken as the norm. The mean was found to increase from one age to the next for all tests except the cancellation test, for which the 9 and 10 year scores are identical. It is impossible to make any direct comparison between our norms and those of other investigators, on account of differences in the form of the tests or the method of giving them.

The norms obtained for each of the nine tests is given in Table I, in the lines marked *Av.* Immediately below each norm, that is, below the average score of each age group, is given the corresponding standard deviation (σ). Our purposes would probably be served by any other measure of variability about as well as by the standard deviation. At the top of Table I, are given the ages of the children and the number (No.) at each age. By age 6.5 is meant all ages from 6 yrs. 0 ms. to 6 yrs. 11 ms. inclusive; similarly for all other ages.

Point Scale Norms for Individual Tests: To deal with the data effectively, it is of course necessary to reduce the norms found for the different tests to comparable units. These units should not only be comparable, but should also be of such a nature that the test which has the highest discriminative value will have its norms represented by the largest number of units, so that, in summing up the total number of units, or points, scored by a child in all tests, the different tests will have weight directly in proportion to their discriminative value.

The discriminative value of a test consists in its capacity for bringing out individual differences in intelligence. This value is low when the scores of children of one age group largely overlap those of adjacent age groups; and it is high when there is little overlapping. Now the amount of overlapping in the scores of two adjacent age groups depends upon

two factors: first, the difference in the average score of the two groups, and second, the variability within those groups.

That the amount of overlapping will vary with difference in the average scores may be readily appreciated. If there is no difference between the average scores of two groups, their scores will substantially coincide, in fact exactly so, except for difference in the form of distribution of the scores. In such a case, if 50 per cent of one of the groups exceeds its own

TABLE I.

NORMS

Tests.	Age	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5
	No.	66	87	82	83	98	95	87	71
Easy Opposites	Av.	0.0	1.1	2.6	4.4	5.2	6.5	7.4	8.6
	σ	0.0	1.2	1.5	2.5	2.1	1.9	2.1	2.6
Hard Opposites	Av.	0.0	0.5	1.3	2.8	3.7	6.3	7.1	9.1
	σ	0.0	1.0	1.5	3.0	3.1	4.4	4.1	4.5
Substitution	Av.	10.1	20.5	30.2	40.3	44.2	50.6	52.3	59.1
	σ	6.2	11.1	11.5	12.0	12.5	11.2	10.7	13.1
Word Building	Av.	0.4	2.3	4.3	6.3	7.4	9.8	10.3	12.0
	σ	0.9	2.5	2.8	3.4	3.2	3.9	3.7	3.0
Digits' Span.	Av.	3.0	4.4	4.9	5.3	5.4	5.7	5.9	6.0
	σ	1.5	0.8	0.7	0.8	0.8	0.8	0.8	1.0
Language Completion	Av.	1.7	5.0	7.5	10.3	11.4	14.2	15.0	15.1
	σ	2.2	3.0	3.1	4.4	3.6	4.7	3.8	4.4
Anagrams	Av.	1.8	4.5	7.4	9.1	9.2	10.2	11.1	11.7
	σ	2.2	3.5	3.1	3.1	3.2	2.8	2.8	3.4
Cancellation	Av.	12.1	16.0	17.2	20.6	20.6	23.0	23.7	25.6
	σ	5.2	4.7	3.9	5.0	4.9	5.1	5.7	5.6
Comprehension	Av.	2.5	4.0	5.8	9.5	13.1	16.1	16.8	17.2
	σ	2.5	2.0	3.9	6.2	5.3	3.9	3.3	2.9

average, we may reasonably expect that substantially 50 per cent of the other will also exceed that average. On the other hand, if there is a great difference between the average scores of the two adjacent groups, there will be little overlapping, and if 50 per cent of both groups exceed their group average, a very high percentage, well over 50, of the older group will exceed the average of the younger group. Obviously, if the variability is the same, the greater the difference in the percentage of the two groups exceeding a fixed degree of ability, the less the two groups will overlap. In general, with a fixed

variability, the amount of overlapping varies with the difference between the averages of the two groups. This difference between adjacent averages, then, is one factor determining the discriminative value of the test.

Equally important is the factor of variability. With a fixed difference between the two averages, the amount of overlapping between the two groups depends upon their variability. We may have, for example, a difference of five units between the two averages, and have either no overlapping or a high degree of overlapping, depending upon whether the children of each group are distributed closely about their own average or spread out over a wide range. If the variability is small, we may find that 100 per cent of the older group exceeds the median performance of the younger; if large, that less than 60 per cent of the older exceeds the median of the younger. As a measure of the variability of any two adjoining age groups, we have used the average of the standard deviations of those two groups.

To find the discriminative value of any test, then, between two adjoining age groups, it is necessary to compare the difference between the average scores of the two groups with the average of their standard deviations. The larger the ratio of the former to the latter, the greater will be the discriminative value. Accordingly we may write, as the formula for calculating discriminative value,

$$\text{D. V.} = \frac{\text{Av}_2 - \text{Av}_1}{\frac{1}{2}(\sigma_1 + \sigma_2)}$$

in which *D. V.* stands for discriminative value, Av_2 and σ_2, respectively for the average and standard deviation of the older group and Av_1 and σ_2, for the same of the younger group.

While the above formula is the one we have used in calculating the discriminative values of our tests, it may aid in understanding their nature to point out that they may be calculated in other ways. This is possible because of the relationship between the standard deviation and the percentage of children exceeding any given deviation from the average. On account of this relationship, discriminative value may be calculated from the percentage of children of one age who pass the average score of those of the next higher age; or from the percentage of the higher age failing to pass the average of the next lower age; or, lastly, from the average of these two percentages. Still another way to put the matter, is to say that discriminative value depends upon the increase with age in the percentage passing a given score.

One way, then, to calculate the discriminative value of a test between two adjacent age groups would be on the basis of the average of two percentages, one, that of the lower group exceeding the average of the higher, the other, that of the higher failing to reach the average of the lower. If the average of these two percentages happened, for example, to be 15.86, then we could conclude that the difference between the two group averages just equalled the average of the standard deviations. This would mean, according to our formula, that the discriminative value of the test in question was 1.0. If the average of the two percentages was anything else than 15.86, we would simply look up the fraction of the standard deviation to which it corresponds. For example, if it was 10, the value of the test in discriminating between the two groups would be 1.28; if it was 35, the discriminative value would be only 0.39.[6]

Again, if we have as data the fact that 75 per cent of the higher group exceed the average for the lower group, then we can, by the use of published tables,[7] read off directly the value of the test in discriminating between the two groups. In the instance given, the discriminative value would be 0.67.

Now it is well known that the percentage of one age group passing the norms of the age group one year younger tends in general to decrease as we reach the higher chronological ages. We should expect, then, that the discriminative values of our tests would show a tendency to decrease at the higher chronological ages. As a matter of fact, they do so.

Before leaving the subject of the relation between discriminative value and the percentage of one age group passing the norms of the preceding age group, it is of interest to consider one conclusion generally conceded to follow from the variation in this percentage. At the higher chronological ages, as already stated, the percentage of any year group passing the norms of the one year younger group is usually less than at the lower chronological ages. This fact, it is commonly recognized,[8] means that the absolute difference between the mental ages of two groups differing by one year is less at the upper chronological ages than at the lower,—that a year of mental age is a smaller unit at the upper ages than at the

[6] See Thorndike, *Mental and Social Measurements*, 1904, p. 148.

[7] Thorndike, *loc. cit.*, or W. F. Sheppard, *Biometrika*, 1903, Vol. II, p. 182.

[8] *Kuhlmann*, Some Results of Examining a Thousand Public School Children with a Revision of the Binet-Simon Tests of Intelligence by Untrained Examiners. *Journ. of Psycho-Asthenics*, Vol. XVIII, 1914, p. 235.

lower. As Kuhlmann puts it, "the amount of mental development during a year cannot be taken as an accurate unit of measurement of intelligence."[9] Mental age is not an absolute unit, but a changing unit. The present method subscribes to this point of view. In fact, it might be summarized as an attempt to find, for the different ages, norms which correspond in magnitude to the actual absolute increase in intelligence with age.

In the present work, the discriminative value of a test as derived from the formula $Av_2 - Av_1 / \frac{1}{2}(\sigma_2 + \sigma_1)$, has been used in preference to that based upon the obtained per cents of one age group falling above or below the norms of adjacent groups. The discriminative values when based upon the standard deviations, as in the formula we have used, are based upon the data for the entire group instead of for portions lying towards the extremes. We have worked through all our data, however, by both methods, and found that in the great majority of instances the results derived by the two methods are so nearly identical as to make it feasible to use whichever is the more convenient. In further work, therefore, while norms are shifting, in order to avoid the labor of calculating the standard deviation anew every time new cases are added, it will probably be advisable to use discriminative values based upon the per cents of the various age groups passing the average of the next higher group and failing to pass the average of the next lower group.

The discriminative values may be treated as "points" or units. Thus, if the discriminative value of a given test between the ages 6.5 and 7.5 is 1.4 units, we consider that the 7.5 year group shows an increase of 1.4 units over the 6.5 year group in the ability tested. We do not, of course, know how many units of ability the 6.5 year group possesses, as we did not extend our testing down to the age of zero, whatever that may be. Taking the average ability of the 6.5 year group as x, however, we may say that the 7.5 year group shows $x + 1.4$ units of ability. Similarly, the quotient representing the value of the test in discriminating between the 7.5 year group and the 8.5 year group indicates the value to be added to the 7.5 year average of ability. Thus, if the discriminative value of the test for the age groups 7.5 and 8.5 is 1.1 units, then we can represent the average 7.5 year ability as $x + 1.4$ units, and the average 8.5 year ability as $x + 1.4 + 1.1$, or $x + 2.5$ units. By continuing this process through all age groups, the original norms are readily changed to norms in terms of points.

[9] *Loc cit.*

In the case of tests of high discriminative value, more points are added for each year, and the norms thus come to be expressed as a larger number of units than in the case of tests of low discriminative value. Because of the different rate of increase in points or units from age to age in the different tests, it results that when we add together the scores for all tests to obtain the final scale, there is an automatic weighting in favor of the more valuable tests. This may be clearly seen, by a comparison of the point scale for easy opposites with that for cancellation. The subject who just equals the 13.5 year norm in easy opposites will secure $x + 4.9$ points, while the subject who equals the 13.5 year norm in the three letter cancellation test will earn only $x + 2.8$ points. This weighting is not done arbitrarily or by any uncertain subjective estimate; nor is it done simply because the opposites test is regarded as a good test and the cancellation a bad one; it is done because the former, by exhibiting at all ages a higher degree of discriminative capacity than the latter, has secured at each age a larger number of points for its norms. How this comes about may be seen from the following table (No. II).

TABLE II.

COMPARISON OF POINT NORMS OF EASY OPPOSITES WITH THREE LETTER CANCELLATION

	EASY OPPOSITES				CANCELLATION			
Age	Original Norms	σ	Discrim. Value	Point Norms	Original Norms	σ	Discrim. Value	Point Norms
6.5	0	0	0.9	x	12.1	5.2	0.8	x
7.5	1.1	1.2		x+0.9	16.0	4.7		x+0.8
8.5	2.6	1.5	1.1	x+2.0	17.2	3.9	0.3	x+1.1
9.5	4.4	2.5	0.9	x+2.9	20.6	5.0	0.8	x+1.9
10.5	5.2	2.1	0.34	x+3.2	20.6	4.9	0.0	x+1.9
11.5	6.5	1.9	0.65	x+3.9	23.0	5.1	0.5	x+2.4
12.5	7.4	2.1	0.40	x+4.3	23.7	5.7	0.1	x+2.5
13.5	8.6	2.6	0.55	x+4.9	25.6	5.6	0.3	x+2.8

The increase with age in the point values of the easy opposites and cancellation norms is shown graphically in Figure No. 1. This figure shows that growth in the ability tested by the easy opposites test is very much greater, when measured in absolute units, than is that in the ability tested by the cancellation test, this in spite of the fact that the original score norms show just the reverse. The increase between seven and thirteen years in the number of cancellations is 13.5, whereas the increase in the number of easy opposites is only 8.6; but the absolute increase in ability with age is nevertheless much greater in the case of the easy opposites.

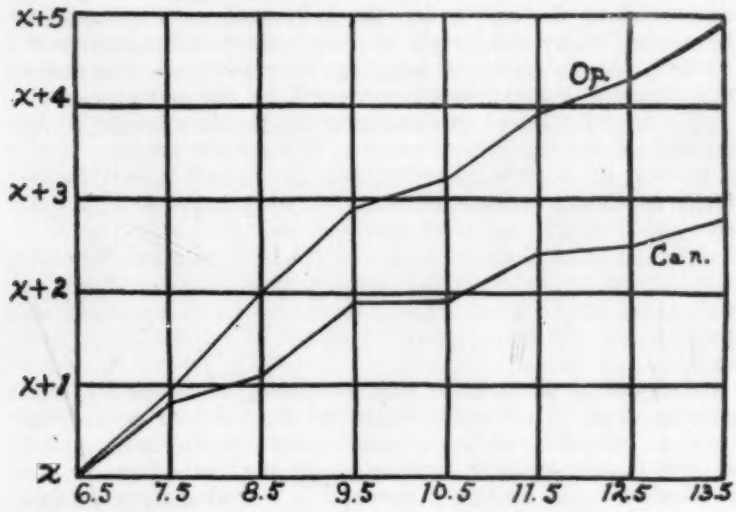


FIGURE 1.— Point norms for two tests, showing weighting effect. Op., opposites test; Can., cancellation test. Chronological ages represented by abscissæ, points by ordinates.

It might appear, at first thought, that in combining the point norms of the individual tests to obtain the final scale, the different tests are not weighted to the same extent toward the x end of the scale as towards the upper end. For example, in Figure 1, since the two curves start at the same point, it may be asked, are they not weighted equally at this point? Such a question, however, would show a misconception of wherein the weighting consists.

To understand this matter, it is necessary carefully to analyze our procedure. What we do, in reality, is simply to combine (by adding their ordinates) a number of curves like those

in Figure 1 to obtain one final curve. Now the *shape* of this final curve is not affected by the height at which we place the individual curves. For example, if we combine the two curves of Figure 1 (by adding their ordinates), the shape of the resulting curve is the same whether the individual curves are plotted so as to coincide at the age of 6.5, as in the figure, or, on the other hand, so as to coincide at the age of 13.5. If we plotted all our individual curves so that they would coincide at the age of 13.5, then x would be the number of units of intelligence possessed by the average child of 13.5, and we would use the scale to measure the intelligence of any individual child in terms of the number of units *below* x. The age at which the curves for the individual tests coincide can be controlled by the height at which the various curves are plotted; but, no matter at what age they are made to coincide, the shape of the combined curve will be the same.

We do not attempt to determine the absolute height of any portion of the intelligence curve. We merely attempt to plot a portion of it—that lying between the ages 6.5 and 13.5—in terms of strictly comparable units. This portion is, as it were, suspended in mid-air at an unknown height. If we knew its height, we would need no x. It does not matter, however, what point we take as the point whose absolute height we call x, for we do know the number of units of height of any point in the obtained portion above or below any other point in it.

It should be clear, then, what weighting a test means in the present case. That test is weighted most whose curve tends most to give the combined curve a vertical direction. If, at certain points, it tends more strongly than at others to give verticality to the combined curve, it is weighted more at those points. On the other hand, that curve which tends most to make the combined curve horizontal is weighted least. With these considerations in mind, it will be clear that the easy opposites test has greater weight from one end to the other than the three letter cancellation test. The point-norm curve for easy opposites, being steeper than that for the cancellation test, tends more to give verticality to any combined curve into which it enters as a component. Further, it will be realized that in those regions where the curve for the cancellation test runs more nearly horizontal than at others, the test has less weight. Thus, between the ages of 9 and 10, where the curve for the cancellation test runs exactly horizontal, the test as a test for discriminating between the mental ages nine and ten has a weight of zero, though it retains a not incon-

siderable weight as a test for discriminating between these ages and those above or below them.

One more question which may arise in connection with the present scale is, why it is called an *absolute* scale? This can best be understood, perhaps, by recalling the facts which cause it to be generally conceded that the absolute increase in mental age from year to year varies, tending to decrease at the upper ages. This generally accepted view rests primarily on the fact that as we reach the higher ages, the percentage of children of one age able to pass the same tests as the average child of the next younger age tends to decrease. It is easy, for example, to find tests passed by only 50 per cent of three year olds, which can yet be passed by 80 per cent of four year olds, whereas it is difficult to find tests passed by 50 per cent of thirteen year olds which are passed by even 60 per cent of fourteen year olds. This variation in the percentage of children of one age who pass the same tests as the average child a year younger is taken to indicate variation in the absolute difference between any two adjacent mental ages.

In short, it is generally conceded that the difference in these percentages serves as an index to the absolute difference in mental ability between any two mental ages. Now the discriminative values which we have used are, as has been pointed out, values which correspond directly with the percentages mentioned; *they are the mathematical measures of that difference between the performances of two adjacent age groups which determines what percentage of the upper group will exceed the average of the lower;* consequently they are the mathematical measure of the absolute difference between the age groups, and a scale of which they constitute the units is an absolute scale.

The original score norms and their values in absolute intelligence units or points is given for all nine tests in Table III. The point values are all worked out in the way already described, illustrated in Table II.

TABLE III

VALUES OF ORIGINAL NORMS WHEN TRANSFORMED INTO ABSOLUTE INTELLIGENCE UNITS

Av.=untransformed norms. Points=value in absolute intelligence units.

Age		6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5
Easy Opposites..	Av.	0.0	1.1	2.6	4.4	5.2	6.5	7.4	8.6
	Points	x	0.9	2.0	2.9	3.2	3.9	4.3	4.9

Hard Opposites..	Av.	0.0	0.5	1.3	2.8	3.7	6.3	7.1	9.1
	Points	x	0.5	1.1	1.8	2.1	2.8	3.0	3.4
Substitution.....	Av.	10.1	20.5	30.2	40.3	44.2	50.6	52.3	59.1
	Points	x	1.2	2.1	3.0	3.3	3.8	4.0	4.6
Word Building.	Av.	0.4	2.3	4.3	6.3	7.4	9.8	10.3	12.0
	Points	x	1.1	1.9	2.5	2.8	3.5	3.6	4.1
Digits' Span....	Av.	3.0	4.8	4.9	5.3	5.4	5.7	5.9	6.0
	Points	x	1.2	1.9	2.4	2.6	3.0	3.2	3.3
Language Completion	Av.	1.7	5.0	7.5	10.3	11.4	14.2	15.0	15.1
	Points	x	1.3	2.1	2.9	3.2	3.9	4.6	4.1
Anagrams.	Av.	1.8	4.5	7.4	9.1	9.2	10.2	11.1	11.7
	Points	x	0.9	1.8	2.3	2.3	2.7	3.0	3.2
Cancellation...	Av.	12.1	16.0	17.2	20.6	20.6	23.0	23.7	25.6
	Points	x	0.8	1.1	1.9	1.9	2.4	2.5	2.8
Comprehension..	Av.	2.5	4.0	5.8	9.5	13.1	16.1	16.8	17.2
	Points	x	0.7	1.3	2.0	2.6	3.3	3.5	3.6
Total.....		**x**	**8.6**	**15.3**	**21.7**	**24.0**	**29.3**	**31.2**	**34.0**

(i. e., absolute intelligence scale.)

Point Values for any Possible Score: After the number of points corresponding to each age norm, as shown in Table III, had been worked out, the next step was to work out the number of points corresponding to any possible score. This was done by the same formula as that used in finding the number of points corresponding to the norms. The difference between a given possible score and the next lower norm was divided by the average of the standard deviations of the two norms between which the score in question occurred. The result, added to the number of points corresponding to the next lower norm, gives the point value of the score in question. Points were assigned in this way for every possible score lying between the norms for 6.5 yrs. and 13.5 yrs. inclusive.

As we had no norms beyond these age limits, the problem of how to deal with scores falling beyond the norms for these ages was a difficult one. It was finally decided to reckon the number of points below x corresponding to any score below the average of the 6.5 year group in terms of the standard deviation of that group; and, similarly, to reckon the number

of points above the 13.5 year point norm corresponding to any score above the 13.5 year average in terms of the standard deviation of the 13.5 year group. Deviations below the 6.5 year norm, then, were transformed into points by dividing them by the standard deviation of the 6.5 year group, while deviations above the 13.5 year norm were divided by the standard deviation for that age group and the resulting number of points was added to the number of points corresponding to the 13.5 year norm. Since x, or arbitrary zero, is placed at the 6.5 year average, the point values of scores below x are negative. They were worked out for each test down to the point value of a performance score of zero.

This procedure used in calculating the point values of scores lying above our highest norm, or below our lowest, differs from the procedure used elsewhere in the scale in that elsewhere the differences are divided by the average of two standard deviations, namely, the standard deviations of the norms both below and above the score in question. Since we have no norms, or averages, below 6.5 or above 13.5, it is necessary with scores respectively below and above these points to use as a divisor simply one standard deviation,— that of the 6.5 year group for scores below the 6.5 year norm and that of the 13.5 group for scores above the 13.5 year norm. Such a procedure involves the assumption that the 6.5 year standard deviation applies to all scores below the 6.5 year norm, and the 13.5 year standard deviation to all scores above the 13.5 year norm. There is an element of error in this assumption, as standard deviations tend to vary with age. It is possible therefore, that by using the 6.5 year standard deviation, we have penalized failure to reach the 6.5 year norm less severely than if the correct standard deviations of the lower age groups could be taken into consideration. For scores above the 13.5 year norm, however, it is not likely that there is any appreciable error involved, since the standard deviations tend not to change much after the 13th year.

We thus worked out a table of nine columns, giving the number of points to be assigned to any possible score in each of the nine tests. Four of these columns are given by way of illustration, in the accompanying table, No. IV. Scores lying outside of the horizontal lines lie outside of the obtained age norms, and were calculated as explained in the preceding paragraphs.

TABLE IV

POINT VALUES FOR ANY POSSIBLE SCORE

Hard Opposites		Language Completion		Comprehension		Anagrams	
Score	Points	Score	Points	Score	Points	Score	Points
0	x	0	—0.8	0	—0.8	0	—0.8
1	0.9	1	—0.3	1	—0.6	1	—0.4
2	1.5	2	0.1	2	—0.2	2	0.1
3	1.9	3	0.5	3	0.2	3	0.4
4	2.1	4	0.9	4	0.7	4	0.8
5	2.4	5	1.3	5	1.0	5	1.1
6	2.7	6	1.6	6	1.3	6	1.4
7	3.0	7	1.9	7	1.5	7	1.7
8	3.2	8	2.2	8	1.7	8	2.0
9	3.4	9	2.5	9	1.9	9	2.3
10	3.6	10	2.8	10	2.1	10	2.6
11	3.8	11	3.1	11	2.3	11	3.0
12	4.0	12	3.3	12	2.4	12	3.3
13	4.2	13	3.6	13	2.6	13	3.6
14	4.4	14	3.8	14	2.8	14	3.9
15	4.6	15	4.1	15	3.0	15	4.2
16	4.8	16	4.3	16	3.2	16	4.5
17	5.0	17	4.5	17	3.6	17	4.8
18	5.2	18	4.7	18	3.9	18	5.1
19	5.4	19	4.9	19	4.2	19	5.4
20	5.6	20	5.1	20	4.5	20	5.7
		21	5.3	21	4.8	21	6.0
		22	5.5	22	5.1		
		23	5.7	23	5.4		
		24	5.9				
		25	6.1				
		26	6.3				

Absolute Intelligence Scale: The quotients which we have termed discriminative values represent an absolute increase in ability from one age to the next, for their respective tests. If the nine tests used can be regarded as tests of intelligence, then a scale built up from an amalgamation of the results of all nine may properly be regarded as a scale of absolute increase in intelligence (beyond a starting point, x, arbitrarily taken as the norm for 6.5 yr. old children). The final absolute intelligence scale is thus secured simply by adding together the point scales of the individual tests. The normal absolute intelligence score for any age is simply the sum of the normal point scores for that age of all nine tests. This has already been given in Table III (bottom line).

The absolute intelligence scale represents the normal growth

in absolute intelligence from the age of 6.5 to 13.5 years. We have plotted this scale as an intelligence growth curve, in Figure 2. It will be observed that the curve brings out the fact that growth in intelligence is considerably more rapid in the earlier years than in the later ones. It looks as if, were it continued at the upper end, it might reach a level somewhere in the neighborhood of 16 or 17 years, and, on the other hand, if continued at the lower end, become much steeper at the first few years of life.

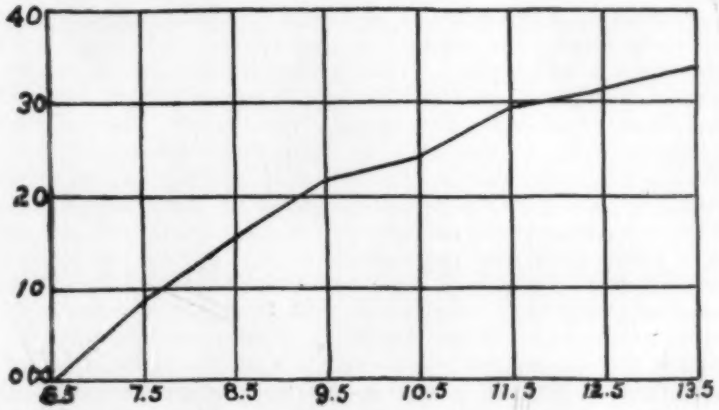


FIGURE 2.— Curve showing growth in intelligence, measured in absolute units, between the ages 6.5 and 13.5. Ages are represented by abscissæ, absolute intelligence units by ordinates.

As already stated, to determine the total point score for each age, the points corresponding to the norms of each age group for the nine tests were added together. Norms for months were then interpolated between these year norms. The final scale, showing the number of absolute intelligence units corresponding to each age, from 6.5 years to 13.5 years, inclusive, by months, is given in Table V.

TABLE V

NUMBER OF ABSOLUTE INTELLIGENCE UNITS CORRESPONDING TO EACH MONTH OF AGE FROM 6.5 YEARS TO 13.5 YEARS

	0 mo.	1 mo.	2 mo.	3 mo.	4 mo.	5 mo.	6 mo.	7 mo.	8 mo.	9 mo.	10 mo.	11 mo.
6 yrs.	....	....	....	....	....	....	x	.7	1.4	2.2	2.9	3.5
7 yrs.	4.3	5.0	5.8	6.5	7.2	7.9	8.6	9.2	9.7	10.3	10.8	11.4
8 yrs.	12.0	12.5	13.1	13.6	14.2	14.8	15.3	15.8	16.4	16.9	17.4	18.0
9 yrs.	18.5	19.0	19.5	20.1	20.6	21.1	21.7	21.9	22.1	22.3	22.5	22.7
10 yrs.	22.8	23.0	23.2	23.4	23.6	23.8	24.0	24.4	24.9	25.3	25.8	26.2
11 yrs.	26.6	27.1	27.5	28.0	28.4	28.8	29.3	29.5	29.6	29.8	29.9	30.1
12 yrs.	30.3	30.4	30.6	30.7	30.9	31.1	31.2	31.4	31.7	31.9	32.1	32.4
13 yrs.	32.6	32.8	33.0	33.3	33.5	33.7	34.0	....	....	....	....	....

By the aid of Table V, mental ages may be determined and intelligence quotients may then be reckoned in the usual manner. The procedure involved, then, in measuring the intelligence of any member of a group of children, by the present method, is as follows: First, the original scores in each of the nine tests is determined. By the use of a nine column table these original scores are changed to point scores. The sum of these nine point scores is the number of units of intelligence credited the child. This number of units can be immediately transformed into mental age by the use of Table V.

Results of Application of the Scale: Measurements of the absolute intelligence above the average for 6.5 years were secured for 490 pupils. These measurements were then converted, by the use of Table V, into terms of mental age and intelligence quotient. It is usually held that the first criterion of the reliability of any scale for the measurement of children's intelligence is that the median child of each age group should be found to have a mental age just equal to his chronological age. Now our norms are based on averages and not on medians, and it is fairly clear that the average mental age of our various chronological age groups runs higher than their median. For most of our groups, however, it is impossible to determine the exact average mental age, for in many cases our only information is that the mental age is over 13.5 or less than 6.5. Consequently, in place of a table comparing average chronological ages with average mental ages, such as we should prefer to give, we must be content with the usual comparison of median ages. By the aid of Table VI, it may be seen how closely the

TABLE VI

MEDIAN MENTAL AGE FOR EACH AGE GROUP

Chronological Age of Group	Median Chronological Age	Median Mental Age
7.5 yrs.	7 yrs. 6.5 mos.	7 yrs. 4.5 mos.
8.5 "	8 " 6.75 "	8 " 4.6 "
9.5 "	9 " 5.4 "	9 " 3.2 "
10.5 "	10 " 6.4 "	10 " 4.5 "
11.5 "	11 " 7.1 "	11 " 5.0 "
12.5 "	12 " 6.2 "	12 " 1.3 "
13.5 "	13 " 5.0 "	13 " 2.5 "

median mental ages parallel the median chronological ages. While the median mental ages run uniformly a month or so below the median chronological ages, they always fall well within the year covered by the ages of the corresponding

chronological groups. So far as this criterion is concerned, we regard our tests as satisfactorily "standardized," though actually we consider them as rather poorly standardized, because we have evidence from the age and school-grade distributions that the different age groups are not equally representative.

As further evidence of the general accuracy of the present scale, we may cite the correspondence between measurements made by it with those made by individual testing. For the purpose of making this comparison, 51 children were tested individually by Kuhlmann's latest revision of the Binet tests. The degree of correspondence is shown, in terms of intelligence quotients, in Table VII. The uncorrected coefficient of

TABLE VII

COMPARISON OF INTELLIGENCE QUOTIENTS OBTAINED BY GROUP TESTING WITH THOSE OBTAINED BY INDIVIDUAL TESTING WITH THE KUHLMANN REVISION.

Chronological Age	Intelligence Quotient		Chronological Age	Intelligence Quotient	
	Group	Individual		Group	Individual
7— 5	113	111	9— 3	114	114
7— 2	106	123	7— 8	121	111
7—10	83	85	7—11	108	116
7— 5	93	98	7— 7	115	117
7— 8	105	121	9— 0	117	124
13— 7	88	99	9— 7	75	75
13— 2	67	73	8— 7	100	102
12— 5	91	82	8— 6	91	102
12— 0	97	98	8— 7	107	117
10—11	79	80	8— 1	94	107
10— 3	79	80	8— 9	93	91
10—10	79	86	8— 8	100	103
10— 7	77	88	8— 3	96	100
10— 4	91	92	8— 0	108	111
10—5	91	93	8— 4	105	101
10—10	103	94	8—10	109	105
10— 6	93	91	8— 6	96	104
10— 7	97	114	8— 1	105	119
10— 8	100	100	8— 9	90	88
10— 1	93	98	8— 2	96	93
10— 0	88	82	8— 7	86	102
10— 2	83	87	7— 1	101	128
10—11	68	77	7— 9	95	103
9— 5	124	124	7—	106	111
9— 0	96	97	7— 7	120	134
9—11	88	92			

correlation between the two sets of quotients is 0.78 (Bravais-Pearson). In view of the greater accuracy of individual tests as compared with group tests, the magnitude of this correlation is rather surprising — particularly so when it is considered that our tests are probably not the best that could be selected and that our norms are not to be regarded as finally perfected. We incline to contribute a considerable portion of the apparent success of the present tests to the method we have used, and to believe that inadequacy of the tests and norms is compensated by the advantages inherent in the methodological principles we have followed.

A further interesting comparison between the results obtained in the present study and those of other investigators is afforded by Table VIII, which gives the distribution of cases

TABLE VIII

DISTRIBUTION OF THE MENTAL AGES OF EACH CHRONOLOGICAL AGE GROUP

Mental Age	Chronological Age 7:0 to 7:11	8:0 to 8:11	9:0 to 9:11	10:0 to 10:11	11:0 to 11:11	12:0 to 12:11	13:0 to 13:11	Total Cases
Less than 6:6	6		1					7
6:6— 6:11	16	1	..	..	..	..	..	17
7:0— 7:11	30	19	6	2	1	..	..	58
8:0— 8:11	17	34	21	11	1	1	2	87
9:0— 9:11	3	13	18	23	11	3	3	73
10:0—10:11	..	3	7	17	18	15	3	63
11:0—11:11	..	1	4	18	14	17	5	59
12:0—12:11	..	..	7	3	8	8	7	33
13:0—13: 5	..	..	1	1	9	5	9	25
More than 13:5	..	..	1	4	19	25	19	68
Total Cases	72	70	66	79	81	74	48	490

by mental age. At the ages of 7.5 and 8.5, the mode and median coincide, and the distribution appears fairly symmetrical. At all of the upper ages, however, with the exception of the small and unrepresentative 13.5 year group, the distribution of mental ages is skewed toward the lower end. Even the 7.5 and 8.5 year groups show a very slightly skewed distribution, when the exact mental ages by months are considered. This is seen, for example, in the fact that the average mental age of the 8.5 year group is 8 yrs. 5.5 mos., while the

median mental age is slightly less, namely, 8 yrs. 4.6 mos. The number of cases is too meagre to justify final conclusions. At present, we may simply note the fact that our data suggest that the distribution of mental ages varies with the chronological age of the group, being almost symmetrical at the ages of 7.5 and 8.5, but becoming rather definitely skewed at the higher ages.

This tendency toward skewed distributions offers an explanation of several minor matters in some of the preceding tables. Since in a skewed distribution it is natural for the median to fall below the average, and since our norms are based on averages, it is possible to understand why, in Table VI, the median mental ages are slightly lower than the median chronological ages. Skewed distributions also account, in part, for the fact, brought out by Table VII, that the present tests very frequently give intelligence quotients which are lower than those obtained by the Binet tests, while at the same time they show, particularly at the upper ages, as seen in Table VIII, an unusually large number of children with a mental age well above their chronological. Pintner, using group tests, likewise found a number of cases showing a mental age greatly above their mental age by the Binet tests, though the median mental age of the group with the group tests was a quarter of a year less than the median with the Binet tests.[10] In a skewed distribution, it is easily possible for a heaping up of cases at the lower mental ages to be accompanied by a considerable number at the more advanced mental ages.

[10] *The Mental Survey*, 1918, pp. 61-63.

"CROSS-OUT" TESTS

WITH SUGGESTIONS AS TO A GROUP SCALE OF THE EMOTIONS

STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF INDIANA UNIVERSITY

By S. L. Pressey and L. W. Pressey

The tests described in the present paper are all "cross-out" tests—that is, each one asks of the subject that by crossing out some one thing he eliminate a wrong, irrelevant, or extreme element in a situation.[1] The writers cannot avoid the feeling that the great stress at present being put upon rigid systematization in method and objectivity in scoring has resulted in not a few tests of form so artificial and unnatural as to seriously interfere with their usefulness. The notion of the elimination of an extra or wrong element was hit upon by the writers as a way in which systematic method and scoring might be obtained, without any such sacrifice of rationale or naturalness in the problem, as presented to the subject. To cross out a mistake is a very natural thing to do. It is a child's own method, as any teacher who has had to train children to erase *instead* of crossing out, can testify. From the subject's point of view the task presented in the tests is thus not at all unreasonable. Such a framing of the problem is, in the second place, satisfactory from the point of view of the examiner, since the directions can be very brief and direct. Finally, so far as scoring is concerned, the plan is admirable; either the wrong word or part is crossed out, or it is not. There is no writing, even of numbers. Indication of the answer by underlining is not wholly ideal (while being at the same time an unnatural and artificial way of answering a question): not infrequently a line is put half way between two words—the writers have found subjects who put the line

[1] The general principle involved in the tests, and certain of the tests described in this paper, have already been mentioned briefly in a previous article, Sex Differences Shown by 2544 School Children on a Group Scale of Intelligence, with Special Reference to Variability, Pressey, L. W., *Journal of Applied Psychology*, Vol. II, 1918, pp. 323-340.

over the right word. But if the line is drawn *through* the word there is a minimum of possibility of misinterpretation.

The writers have now tried this principle of the "extra" or "wrong" element in a variety of tests, with a variety of subjects, and have found it to work excellently. The present paper will describe briefly (1) a brief group scale of intelligence, to be used in school surveys, and applicable from the fourth through the eighth grade, and (2) a scale supplementary to this, for use from the kindergarten through the third grade; (3) a group scale for measuring emotional interests will be outlined, and suggestions made as to a test of learning, a test of ingenuity with geometrical forms, and a silent reading test. There is here surely variety enough; but all these tests make use of this same "cross-out" principle. And one might almost say that a single brief direction would serve for all: "Cross out what is wrong, or does not belong with the others."

I. Survey Scale Unit E

The blank is partly reproduced below, only the first 5 of the 25 items in each test being given here. Briefly the directions are as follows: in the first test the children are told that each list of words is a sentence with the words mixed up. But there is in each list one word that does not belong in the sentence. This word they are to cross out. In the second test the children are told simply to cross out in each list the thing that does not belong with the other four things. For the third test it is explained that the numbers in each list are arranged in accordance with some rule, but there is one number in each list that breaks the rule; and this recalcitrant number they are to cross out. Finally, in the last test the children are told very simply to "Cross out in each list the thing that is worst." Five minutes are allowed for the first and third tests, four minutes for the second and fourth.

The results below are from 850 school children (grades four to twelve inclusive) in an Indiana city of about 12,000 population. By total score on the four tests the results are as follows:

Grades	4	5	6	7	8	9	10	11	12
Number	178	147	111	89	68	102	56	42	57
Medians	40.5	48.1	56.1	60.0	69.5	71.1	72.0	74.0	75.5
25%	32.7	43.1	48.1	55.1	64.0	66.4	67.5	64.2	72.6
75%	47.2	54.2	60.2	68.2	75.0	74.6	77.2	81.0	80.0

Results by the individual tests were for each grade as given in Table I. It will be seen that there is a fairly good progress

Schedule E.

MENTAL SURVEY SCALES

"CROSS-OUT" TESTS

Name ..... Sex .....
Age ..... Birthday ..... Grade .....
Teacher ..... School .....
Place ..... State ..... Date .....

TEST I. VERBAL INGENUITY

Examples: (a) see a I man on.
(b) knife chair the sharp is.
(c) cat dog the house after ran the.

(1) the cat at see.
(2) boy was sky the sick.
(3) bread sweep will the kitchen I.
(4) are going yesterday to-morrow we.
(5) me mine give my straw hat, etc.

TEST II. LOGICAL JUDGMENT

Examples: (a) dog, cow, horse, oak, cat.
(b) book, newspaper, magazine poster, photograph.
(c) chair, desk, table, wagon, book-case.

(1) coat, shoes, hat, gloves, cow.
(2) oats, wheat, barley, cotton, rye.
(3) satin, silk, linoleum, calico, serge.
(4) fly, ant, bee, grasshopper, mouse.
(5) bread, meat vegetables, hay, fish, etc.

TEST III. ARITHMETICAL INGENUITY

Examples: (a) 2 4 6 8 9 10 12
(b) 7 6 5 6 2 7
(c) 1 4 7 9 10 13

(1) 1 2 3 9 4 5
(2) 2 4 6 7 8
(3) 19 18 17 16 13 15
(4) 5 7 10 15 20 25
(5) 11 10 8 6 4 2 etc.

TEST IV. MORAL JUDGMENT

Examples: (a) Gambling, lying, drunkenness, murder, smoking.
(b) begging, gambling, trading, cheating, borrowing.
(c) dullness, foolishness, laziness, weakness, poverty.

(1) fighting, killing, hating, quarreling, hurting.
(2) borrowing, gambling, overcharging, stealing, begging.
(3) love, hate, fondness, dislike, liking.
(4) dancing, drunkenness, flirting, over-eating, smoking.
(5) holiness, reverence, piety, obedience, wickedness, etc.

from grade to grade through the eighth (as far as the tests are expected to apply), with a level after that.

As regards the structure of these tests the writers wish to call attention to the following points: (1) the possibility of chance successes is at a minimum; in the second and fourth tests the chance is one in five, in the first and third, still less; (2) there is possible great condensation in the printing of the tests; an examination of one hundred clear-cut problems is presented on two sides of a single 12 in. by 9 in. sheet of paper. Such condensation is saving of space, trouble in scoring and handling, and expense of paper and printing; (3) the method by which the subject indicates his answer is the extreme of simplicity; he simply crosses out a word. The manual labor of taking the examination is thus at a minimum; and the rate at which the subject progresses is determined solely by his ability to solve the problems. Finally (4) scoring is extremely easy, rapid, and wholly objective. There is in each line only one correct answer. The subject has indicated his response to each problem not by writing, nor even by so simple a thing as a number, but by drawing a line through a certain word or number, one only in each line. It may be said that the examination takes less than 25 minutes to give, and the blanks can be scored at the rate of 70 an hour by student help.

SURVEY SCALE UNIT E

RESULTS BY TEST

Grade	Test I.			Test II.			Test III.			Test IV.			Grade
	25%	M.	75%	25%	M.	75%	25%	M.	75%	25%	M.	75%	
4	8	10.7	14	10	11.9	14	7	9.7	13	7	9.4	11	4
5	11	13.6	16	11	13.0	15	9	12.1	14	10	11.6	13	5
6	13	16.6	18	12	14.0	16	11	13.1	16	11	13.0	15	6
7	15	17.7	20	14	16.1	18	12	14.9	17	13	15.4	17	7
8	17	19.5	21	17	18.7	20	14	16.5	18	16	17.3	19	8
9	18	20.1	22	17	18.5	21	14	16.2	18	15	17.3	18	9
10	18	21.5	23	18	19.6	21	15	16.7	18	16	18.2	20	10
11	16	20.0	23	17	19.6	21	14	16.8	18	17	19.4	21	11
12	19	21.1	23	18	19.4	21	15	16.9	18	18	20.5	22	12

But the writers wish to emphasize particularly certain features of the examination as it appears from the point of view of the subject. There is to each test a rationale and a plausibility that immediately arouse both interest in and understanding of the test, with only brief directions. The subject's response—the crossing out of an extra, incongruous, or undesirable element—is so natural as to be considered a matter of course. Further, the directions are in essence the same for all four tests, and the subject's response is the same. It may be said that for 850 subjects, from the fourth grade on, there

were obtained no zero scores on the total examination. There were only 8 zero scores on separate tests (5 of these on the third) and only 20 scores below 3 on any test (9 of these on the third).

The writers have been quite as much interested in the individual tests as in any ratings that might be obtained from the total examination. The first test might be expected to give a measure of language ability. As such it should have definite superiority over the Trabue Completion Scales in its absolute objectivity of scoring and its elimination of writing. It should in addition sample more complex mental processes, since the mental reconstruction of a sentence would seem a more elaborate effort than the mere completion of it. It should also be pointed out in passing that this form of "dissected sentences" is greatly superior to the form used in the army scale Alpha in its condensation of matter, its elimination of the possibility of chance success, and its elimination also of elements of information from the problem.

The second might be called a test of information with understanding.[2] It should be noticed that the test presents in each line a problem in both differences *and* similarities. The subject must make sure that four things in each list are alike, and the fifth different from *all* the other four. The Binet separates the problems and asks either "how are these two things different" or "how are these two things alike." Group test phrasings of this problem have been of two forms; in one the subject is presented with pairs of similarities or opposites, and is asked to underline either the "similar" or "opposite" as one or the other applies; in the other he is given a word, followed by a number of other words from which the opposite or similar is to be selected by underlining. The superiority of the present phrasing in compactness, naturalness and simplicity of direction would seem unmistakable.

The writers have hoped that the "arithmetical ingenuity" test might give a measure of potential arithmetical ability, and so give results in interesting contrast to results with tests of the fundamental processes in arithmetic, in classes where the teaching was poor but the abilities of the children good. The

[2] A test of this form for technical knowledge in a particular field would seem possible. A skilled mechanic might be given lists of various tools or processes, with in each list one process or tool not properly belonging there; four lathe parts and a milling machine part, for instance.

test in this present form, however, is not wholly satisfactory.[3] The product moments correlation for this test with results from a group scale of determined reliability previously described in this JOURNAL[4] shows for 205 6th, 7th and 8th grade children a coefficient of only .34 as compared with .61 on the first test, .65 for the second, and .91 for the last.

This high correlation of the last test with intelligence ratings otherwise obtained is extremely interesting. The test undoubtedly involves vocabulary, and general capacity for abstract thinking, as well as moral sentiment and judgment. In formulating the test the writers had in mind, of course, these last abilities especially, and they hope to obtain results on this test from the state reform school, to see whether any relation appears between ethical discrimination, as this test measures it, and delinquency. But altogether aside from the interest of its special content, the test has, apparently, a surprisingly high value as a measure of "general ability."

Mental Survey Unit E is thus a scale of four tests, each test of independent interest, the combined result giving also something of a rating of "general ability." The writers have thought of the scale as of value chiefly for the purpose of comparing classes, schools, communities, or other groups. As they have shown elsewhere,[5] such comparisons may be of distinct value in the study of educational problems. They hope to present shortly data bearing on the sociological value of such group measurements. And they feel that for such investigation "research units" of the general nature and form of this small four-page examination, taking approximately 20 minutes to give and scored with extreme readiness, are especially satisfactory.

II. SURVEY SCALE UNIT F

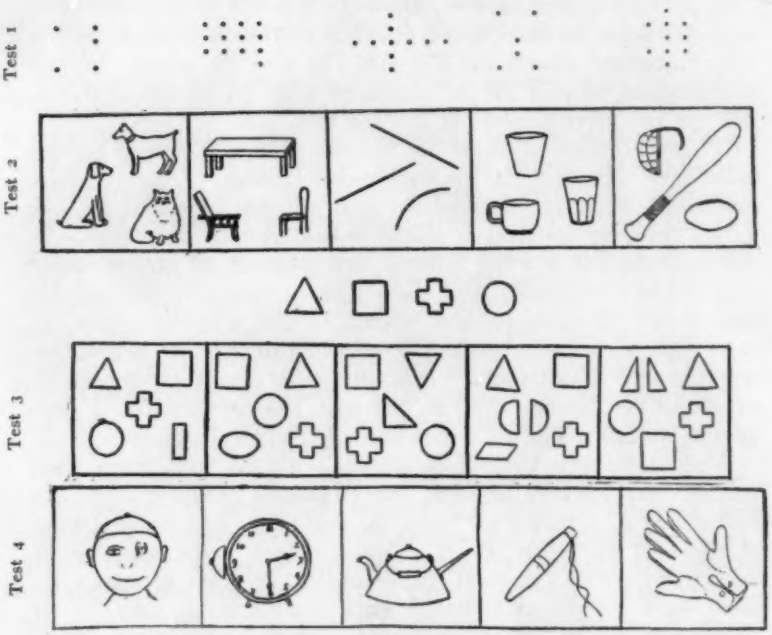
Samples of the four tests of the examination are shown below:

[3] The test is very poorly graded, and involves too great a variety of problems. It is now being revised so that the "numbers count up or down, multiply or divide in some regular way;" such irregular progressions as are shown in the thirteenth item are dropped altogether.

[4] A Group Point Scale for Measuring General Intelligence. *Journal of Applied Psychology*, Vol. II, 1918, pp. 250-269.

[5] Pressey, S. L. A Comparison of Two School Systems by Means of a Group Scale of Intelligence, *Educational Administration and Supervision*, Vol. V, 1919, pp. 53-62.

PLATE I



In the first test the children are asked to cross out the extra dot; in the second to cross out the thing that doesn't belong with the other two; in the third to cross out the block left over, in each square, after the other blocks have been fitted into the four forms (square, triangle, cross and circle) at the top of the page; in the fourth test to cross out, in each picture, the part that is wrong. Each test contains 25 items (with 5 items additional for examples). The children are allowed three minutes for each test. Results from 470 children in the 1st, 2nd, and 3rd grades of the same system surveyed by means of scale E show the following results:

Grades	1	2	3
Number	165	158	147
Medians	34.7	48.3	57.0

The medians for the separate tests at each grade were as below:

Grade	Number	Tests I	II	III	IV
1	165	5.5	10.9	6.4	13.3
2	158	10.4	12.4	10.6	17.0
3	147	13.3	13.6	14.0	18.4

It will be seen that the tests are not as yet properly graded, but the general principle of each one of the four tests has proved satisfactory. All four tests can be "gotten over" even to kindergarten children, and indications to date would lead one to expect a high correlation with intelligence as soon as the problems are properly graded.

The problem faced, in building these tests, was not an easy one. It was desired (1) that the tests should not require literacy, not even so much as knowledge of the alphabet or the cardinal numbers; otherwise more or less incidental influences such as early home teaching and training might play a disproportionate part in the results. The difficulty was to present *without* the use of any such symbols, problems of such a complexity as would give a measure of general intellectual ability. It was also desirable (2) that the children should indicate their answers in some extremely simple and unmistakable way. Young children show too poor eye-hand coordination to ask them to complete a picture, follow through a maze, or go through similar complicated processes, in the group, without greatly lengthening the examination, wearying the children, and introducing difficulties of scoring which would make the examination almost impracticable. Finally (3) it became evident very early in the experimentation that very young children are as yet so little disciplined in habits of attention that directions of any length, or problems of any artificiality or lacking in intrinsic interest, were not to be thought of. The directions must be almost colloquial in their simplicity and directness; the test must, in fact, be self-explanatory. And the examination must have more the appearance of a game, than of a task to which there must be conscious application.

The "cross-out" idea seemed admirably suited to the problem. The dot pattern test was tried with some hesitation. The results show the items to be poorly graded, but the problem of the test was better understood by the child than was expected (and this in a system without any kindergarten training which might afford some special preparation for such a test). So far as could be judged from the data now in hand, there is a fair correlation with general ability.

Test II involves the same problem as Test II of the previous scale (in picture form) and needs no further comment. Test III is an attempt to put a form board test on paper, and in such a way as could be used in the group. It is the result of a long series of experiments and has in it possibilities of further development which will be returned to shortly. The last test, which is obviously the best of the four technically, permits the presentation of an indefinite variety of problems, and of a range of hardness from 98% success in the kindergarten to problems difficult enough to puzzle an intelligent adult.[6]

But the interest in the Primer scale has been not so much in the separate tests as in the total rating. The scale is an attempt to make possible a survey of the *entire* pupil material of a school or school system. As a matter of actual fact the data thus far presented is the result of such a survey of *all* the school children, *from the first grade through High School,* in a city of about 12,000 population.[7]

The writers have already[8] presented data to show that schools may differ markedly as regards the grouping which they make of the children, as regards ability, in grades and sections. A survey which shall include the *entire* school system is thus necessary if an adequate picture is to be obtained of the school's disposal of its pupils. And a measure of the abilities of the children before such selective and formative influences have had much opportunity to operate is necessary, if a fair estimate of the community is to be obtained. This primer scale should then, if found satisfactory, give results of the greatest sociological as well as educational significance. In fact, for all such group comparisons it should give data of unique interest.[9]

---

[6] The test has interesting possibilities, as a trade test for example. A picture showing a piece of machinery with some part out of order—this part to be crossed out—would be excellent for such a purpose.

[7] It might be added that the total 1,321 children were tested in a day and a half, the primer scale being given exclusively by one of the writers and Scale E throughout by the other, and that the entire 1,321 blanks were scored by student help at a total cost of $8.00.

[8] See note 5 above.

[9] It may be said, shortly, that children entering the schools located in the best neighborhood of the city surveyed averaged over a year in "mental age" above children (of the same chronological age) entering another school in a poorer section of the city, and that the median rating of grade 1 in the best school was approximately the same as the rating of grade 2 in the poorest.

III. Suggestions as to Other Tests, Especially of the Emotions

The following tests have been only superficially experimented with, with small groups of school children or college students. There is thus no systematic data to present. But the tests are of some intrinsic interest. The writers are hardly likely to find time for further work with them for some time to come; and since they are based on the same "cross-out" principle, they are described briefly here.

No one can have had any intimate experience with the feebleminded or the insane without feeling very strongly at times the almost pitiful inadequacy of a test of "general intelligence" as a single means of measuring social and economic inadequacy. One very important element in such cases is lack of emotional control. The writers wish to suggest the following four tests as making an interesting research unit for experimentation in the field.

1. The first test (to be placed on the first page of a four-page folder) consists of a story, beginning in very monotonous and colorless fashion, but in the middle of the page developing into a hyper-adventurous train robbery episode of the dime novel type. Sown throughout the test are irrelevant words, 20 in the first section and 20 in the second. The directions are to read the passage as quickly as possible, crossing out all the irrelevant words. The score consists of the difference between the number of words correctly crossed out in the exciting, as compared with the uninteresting section.

2. The second is a group adaptation of the familiar association test. There are 20 brief paragraphs of pied type on the blank. The subjects are told that 20 words will be read to them; that as soon as the first word is named they are to cross out all the letters in the first paragraph which occur in this word, and so on through. The first ten words are not emotionally toned; the last ten, such as are likely to touch a "complex" or stir up some emotion. And the score is the difference between the letters correctly crossed out in the two groups.

3. The third test consists of twenty-five lists of words, five words in each list. Some of these words are, for the average person, highly emotionally toned, some slightly and some not at all affectively colored. The subjects are told to go through the lists and cross out *all* the words that are unpleasant. After this is done, they are told to go through the lists again and, in each list, draw a line around the word that is *most* un-

pleasant. The slight experimentation with this test would suggest that the number of crossed words gave something of a measure of a person's tendency to emotionalize — or of affective spread. It is hoped that the number of deviations, in the selection of the most unpleasant word in each line, from the most common selection, may give a measure of neurotic "transvaluation."

4. The last test consists of a list of about 150 words. Of these half occur in the previous three tests, half do not. Of the words occurring in the previous tests, half are words with strong emotional coloring and half are words with no coloring. And of the words in the list, not appearing in the previous tests, half are emotionally toned. The subjects are told to go through the list and cross out the words that they have seen before, in the other tests. The record is scored as to (a) the proportion of emotional to unemotional words remembered and (b) the proportion of emotional to unemotional words remembered wrongly.

It has been the hope of the writers that they might find time to work out this four test unit with some care and obtain for purposes of comparison data from school children, a group of institutional feebleminded, and representative groups from the boys' and girls' state reform schools. No such results have been obtained as yet. But the writers are convinced, after a somewhat unusually varied and intensive experience in work with both the insane and psychopathic and the feebleminded, that in such extreme cases, at least, emotional instability is a gross enough thing to be measurable by some such means. And they feel very strongly that a study of the problem by *group* comparison is particularly needed. Individual differences thus drop out and it is possible to get at type reactions.[10]

When there is any considerable foreign element in a community tests intimately involving language can hardly be adequate. Other material is then necessary (such material may very well prove indispensable in all cases). That is, a group performance scale supplementary to scale unit E is needed. As two tests for such a scale the writers have experimented with (1) a puzzle test similar to the form board test above, showing at the top of the page a large square, and below in lines, groups of blocks, those in each group just sufficient to

[10] In fact they have an idea that such tests might get at a person's emotional peculiarities better when that person felt himself sheltered by being submerged in a group than would be the case in an individual examination.

fill the square but with one block left over. The extra block is, of course, crossed out. The writers have already obtained results from about 200 children and find the test highly satisfactory. As a second test (2) the writers have already in trial form a variation of the digit-symbol test. It consists of the usual blank, but with the substitutions already made, and in each line one substitution made wrongly. The children are told that in each line there is one wrong substitution, and they are to cross it out. The saving of time and labor in taking the test, and of trouble in scoring, over the standard form is obvious.

Finally the writers wish to suggest a new form of silent reading test employing the cross-out principle. Silent reading tests seem to fall into one of two pitfalls, either the pupils' performance cannot be objectively and satisfactorily scored or else his performance is more or less artificial, and with only a somewhat indirect bearing on the passage read. The writers have experimented tentatively with a silent reading test having such items as: "In winter the wind is very cold; it blows against our house, and often finds its way in through the cracks and keeps the house warm." The directions are to cross out the one word in each sentence that spoils the meaning. The problem as thus phrased is easily presented to the pupils, the scoring is wholly objective, and labor in both giving and scoring of the test is slight. The problem presented seems, to the writers at least, much less artificial than is usually employed in such tests and much more truly a test of comprehension of the meaning of the passage read.

IV. Group Tests as a General Method of Research in Psychology

A word remains to be said in defense of the presentation, in a single paper, of such heterogeneous material, much of it still in only experimental form. The writers have brought these various bits of experimentation together in just this way because they wish to emphasize the almost bewildering variety, in material, in method, and in problem, possible with the "group test." The tentative suggestions regarding four tests of the emotions will very likely be taken by the reader as amounting to little more than an opportunity for parlor amusement. For the writers these few paragraphs are perhaps the most serious part of the paper. They represent an attempt to formulate—in group test form—certain materials for dealing with one of the most baffling problems of psychology. The

"tests" are not intended as a means of diagnosis, but rather as a method of investigation. The writers believe that the majority of psychological problems can be attacked with profit by the "group test" method. They feel that it would be very unfortunate if the methods brought to general notice by the army psychological work, of investigation in the group, of systematic condensation of matter and record, and studied economy of time and labor in giving and scoring, should be limited in their application to the study of "general intelligence." Particularly are studies of groups *as such* likely to prove fruitful; we should thus be able to determine types, and to relate psychological study to large educational, sociological, and economic facts. The "group test" has appeared in the army work as a new method for dealing with certain problems of general intelligence; in reality it is a new method of investigation, applicable to the entire field of psychology.

Summary

The paper may be briefly summarized as follows:

1. The present emphasis upon rigid standardization in method and objectivity in scoring is tending toward tests of form so artificial and unnatural as seriously to interfere with their usefulness. As one way of obtaining greater naturalness—without sacrificing method—the writers suggest the formulation of test problems in such a way as to involve the elimination, by crossing out, of a wrong or unnecessary element.

2. First forms are presented of two brief group scales making use of this device; one for grades one to four inclusive, the other for grades four to eight,—the two together permitting the survey, for purposes of group comparison, of entire schools and school systems, from the kindergarten up.

3. Certain tests of special abilities, and a brief scale designed to measure emotional control, are suggested.

4. A plea is made for the recognition of the "group test" as a general method of investigation in psychology.

A COMPARATIVE STUDY OF THE VARIABILITY OF BOYS AND GIRLS

By GEORGE W. FRASIER, Washington State Normal School, Cheney.

During the past decade much has been written and said about the variability of the sexes. One school of psychology maintains that there are no median sex differences of mental capacities, but the ability of the male is scattered over a greater range, that is, he is more variable. They thus account for the fact that more men than women have become famous, also that more men are in the low levels of society. The other school of psychology believes that there are no sex differences either in variability or central tendency. Furthermore they maintain that it is the present constitution of society that puts men into the extremes. The question of variability as applied to children is of vast practical importance to those who deal with boys and girls. Shall we deal with children on the assumption that boys *should* be found in the extremes and the girls in the middle ground?

The literature of the subject can be separated into two classes (1) studies dealing with men and women and making no mention of boys and girls and (2) those dealing with children. It is evident that many students of this subject draw their data from the behavior of adults, make their conclusions on such data and then apply their findings in dealing with children. This attitude is not justified by facts. The following study deals with school children, and for this reason the quotations taken from other authors will be confined to those dealing with boys and girls.

Keyes made a special study of all the children in one school district, having six or more annual records available. He found 1,902 children distributed as is shown in Table I.

TABLE 1

	Boys	Girls
Repeaters	373	310
Gainers	314	299
Normal	302	304
Total	989	913

He concludes from the above table [1] "The boys in this group (repeaters) are more numerous than the girls, as would be expected from the known greater variability of the male," also "The gainers or accelerates were 613 in number; 314 of them boys, 299 girls. The greater variability of boys is here again shown."

A casual glance at the table seems to vindicate the conclusion made by Keyes. However, it is to be noted that the totals show 76 fewer girls than boys. Now, Keyes has assumed that had these 76 girls been present all of them would have been in the "Normal" group. To overcome this inequality and show the true comparison between the boys and girls it is necessary to reduce each group to the percentage of the total. The percentage grouping is shown in table 2.

TABLE 2

	Boys	Girls
Repeaters	37.7%	33.9%
Gainers	31.6%	32.7%
Normal	30.5%	33.3%

The above table indicates that when these data are reduced to a comparable basis the greater variability of the boys vanishes.

Thorndike believes that boys are more variable than girls, for he writes [2] "Athough the male and female types are closely alike in intellectual capacities, there is an important difference in the deviations from the type in the two cases, namely, that the males deviate more. The highest males in any quality are more gifted than any of the women, and the lowest males inferior to all women. Thus, though girls in general rank as high or higher than boys in high school and college, they less often lead the class; thus there are far more eminent intellects among men than among women and twice as many idiots." In a foot-note he offers the following explanation: "This difference is not absolutely proven to exist but it appears fairly certain from the results of several investigations."

Rogers and McIntire in measuring the intelligence of children with the Binet-Simon scale found that [3] "The principal outcome is the familiar law that there is a wider scattering of ability among boys than among girls; consequently there are more gifted as well as more backward boys than girls," and

[1] C. H. Keyes. *Progress through the Grades of City Schools*, p. 15, 17 and 32.
[2] E. L. Thorndike. *Principles of Teaching*, p. 96.
[3] *British Journal of Psychology*, 1914, vol. 7, p. 265.

"so far as we go, our experiments support the general law, which several observers have already noted, that there is a greater variability of capacities in boys than in girls, but no general superiority."

The most complete study made in America dealing with the case in point was made by Terman. Terman studied 1,000 unselected school children from the standpoint of intelligence and came to the conclusion that [4] "The supposed wider variation of boys is not found. Girls do not group themselves about the median more closely than do boys. The range of I Q including the middle fifty per cent is approximately the same for the two sexes."

Leta Hollingworth made a very complete resumé of studies concerned with the variability of the sexes and pointed out that up to 1914 these had not established the greater variability claimed for the males. She says "—and there is at the present time no conclusive empirical evidence to show that in cases where the coefficient of variation is greater for one sex than for the other, this greater variability consists of greater range." [5]

It appears from the quotations given that the subject is far from settled, and for this reason the following data are presented.

It is claimed by the school of variability that whenever males and females, or boys and girls, undertake a task that the best progress and the slowest progress will be made by the boys, and the girls will compose the great bulk of median ability. Then, one would expect this law to hold true in school accomplishment.

It appears that should we select all of the boys and girls of a certain age in any school system or systems we could, by finding their grade location, determine the relative variability of boys and girls of that age.

Thirteen was selected as a desirable age because (1) children of this age are held in school by the attendance laws in most states and (2) at 13 a child has been in school long enough to demonstrate his ability.

The grade location of all thirteen-year-old boys and girls was found for twenty cities:[6] Boston, Mass.; Columbus, Ohio; Wichita, Kansas; Grand Rapids, Mich.; Kansas City, Mo.; Richmond, Va.; Winston-Salem, N. C.; Baltimore, Md.;

[4] L. M. Terman. *The Measurement of Intelligence*, p. 70.
[5] *American Journal of Sociology*, 1914., vol. 19, p. 510-530.
[6] The data concerning these children were taken from the city school reports.

Cleveland, Ohio; Trenton, N. J.; Philadelphia, Pa.; Danbury, Conn.; Fall River, Mass.; Winchester, Mass.; Beverly, Mass.; York, Pa.; Williamsport, Pa.; Aurora, Ill.; St. Louis, Mo.; Waterloo, Iowa. Table 3 gives the total number of each sex in each grade.

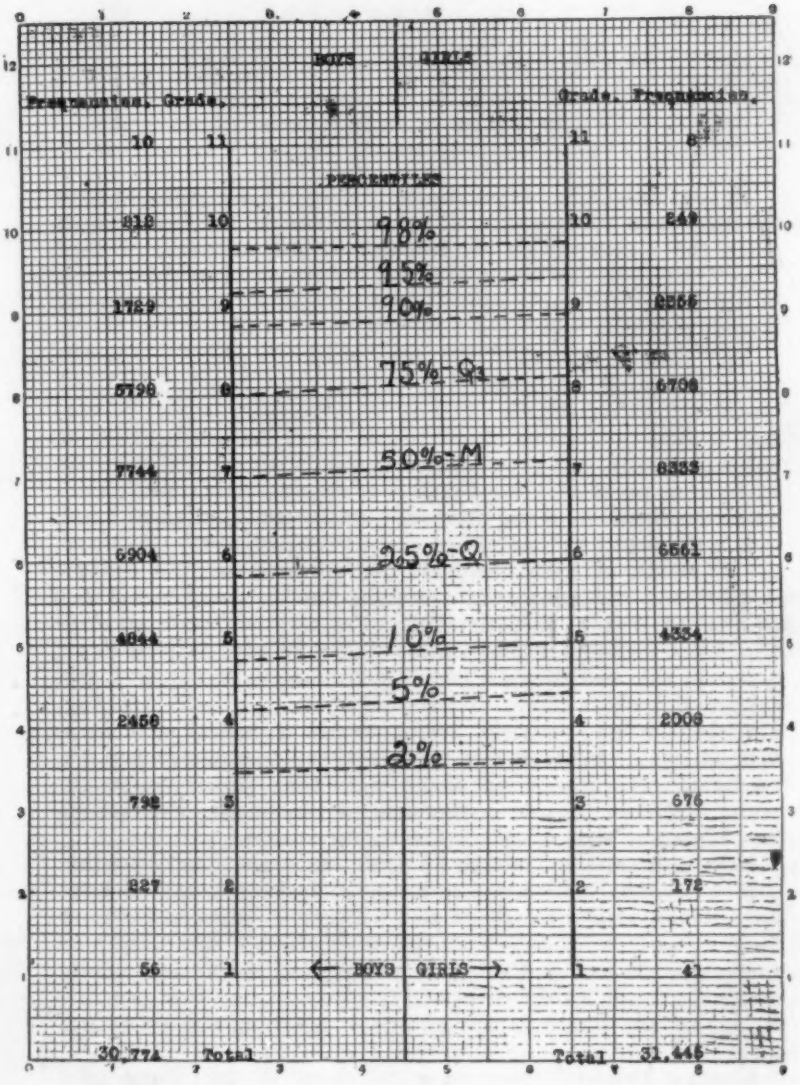
TABLE 3

Grade	Boys	Girls
I	56	41
II	227	172
III	792	676
IV	2,458	2,008
V	4,844	4,334
VI	6,904	6,561
VII	7,744	8,333
VIII	5,798	6,708
IX	1,729	2,355
X	212	249
XI	10	8
Total	30,774	31,445
Median	7.01	7.24
Q^1	5.81	6.10
Q^3	8.01	8.22
Q	2.2	2.12
C. V.	.16	.14

The coefficient of variability for the boys is slightly greater than that for the girls but the difference is so very small that no greater variability can be said to exist. In order to show all possible comparisons between the boys and girls chart A is given showing nine different percentiles, for each sex and their relative locations.

It is evident from the chart presented that girls make better school progress than boys. From the standpoint of variability our data point clearly to the conclusion that no sex differences are shown. The statistical procedure is carried out from the second to the ninety-eighth percentile so that the question of greater range is disposed of.

It is safe to conclude from the study of 62,219 thirteen-year-old boys and girls in 20 cities that the greater variability claimed for the boys is not present. It appears that this greater variability has never been established with any large number of cases or with any degree of certainty and if it is to be used as a working basis for those dealing with boys and girls it will be necessary for its supporters to present some conclusive empirical evidence. Until this is done it seems safest to disregard such a theory and deal with boys and girls on the supposition that they are alike so far as mental variability is concerned.


BOYS
GIRLS
Progressive Grade.
Grade. Progressive.
PERCENTILES
99%
95%
90%
75% Q3
50% M
25% Q1
10%
5%
2%
← BOYS GIRLS →
Total
Total

MENTAL ATTITUDES IN DEBATE

By Walter B. Swift, Boston

In speaking of mental attitudes in debate I am going to limit the subject to the moments of actual debate, to the few minutes when the speaker is on his feet before the audience. I shall have nothing to say about the time before or after the speech. What I have to say concerning development and change of mental attitudes will come later.

By "mental attitude" I mean "state of mind." Another name for mental attitude is "feeling," and another is "disposition." However, this feeling or disposition is not the normal feeling or disposition of the debater, the feeling that he holds generally, but is one which he holds during the moments when he is presenting his debate. It dominates him and dominates the situation at the time, but, unlike usual states of mind, it is not caused by the situation. On the contrary, it causes the situation, of itself. Again, the words which the debater is using do not cause the situation, but it dominates those words and gives them their effect. For example, suppose some one arises and says a few words. What he says may be clear, grammatical, and thoughtful. Yet the words themselves will not tell the whole story. There will be behind them a feeling, a state of mind — what I prefer to call a mental attitude — that will show through the words and impress itself upon the listeners. The words will mean more than they actually express, and they will do so because of the feeling that creeps into them and through the voice of the speaker.

Now as a rule, in the club of young debaters, most attention is given to the English and to an oratorical utterance. I have no fault to find with either sort of thing — indeed, I realize their value and I favor and encourage them. But I wish to emphasize the fact that this side of debating, which is a wholly external one, is not the only side to consider, and that possibly the other, the internal, mental one is after all the more important of the two.

Keeping this difference in mind, then, let me illustrate a little.

The usual way of reciting Yankee Doodle is somewhat like this: (Simple)

Yankee Doodle came to town
Riding on a pony.
He stuck a feather in his hat
And called him macaroni.

I can now so change my state of mind that an entirely different presentation is called for: (Sad)

Yankee Doodle came to town
Riding on a pony.
He stuck a feather in his hat
And called him macaroni.

Here let me call attention to one important thing: that each time I recited the stanza *the words were exactly the same.* Whatever change, therefore, was noticed in the reading was due to something else, to the change in my mental attitude as I recited the lines.

Here one may object that these two ways of reading are the only ones possible to this particular bit of verse. Let me answer by a third reading: (Revengeful)

Yankee Doodle came to town
Riding on a pony.
He stuck a feather in his hat
And called him macaroni.

These three ways of reciting the same words are sufficient to show that the change in the mental background may be a very varied thing, and that as the mental attitude changes, the rendering of the spoken word changes too.

Now what happened in my mind to cause this difference in my three readings? It was not a mere trick of the voice, but a real change that occurred much farther back than in the vocal mechanism, a change in the working of the mind that governs the vocal cords.

When I read Yankee Doodle the first time I did so in the usual way, giving the words the natural interpretation. But the second time I caused myself to think that Yankee Doodle's mother had just died; that he had come to town to attend her funeral; that he and his mother had been intensely dear to each other; and that her death was in consequence a great loss to him. This gave the recital an atmosphere of intense sadness.

At the third reading I pictured to myself the disastrous dealings of an enemy and the return of Yankee Doodle to the town for revenge. He had come to punish underhanded doings. Thus the reading reflected the change in mental attitude.

Just what I mean by mental attitude should now be clear. It is not personality, it is not quality of character, it is not a mental trait; it is rather an interpretation of the situation, it is the color or imaginative turn that one gives to his mind under the stimulus of a situation; it is the way that one deliberately makes himself feel in order that he may give force and an arousing power to his words.

Some may think that this is a little thing, that it is limited to little feelings or little mental attitudes which anybody can assume at will. I am going, therefore, to give two or three brief examples, in order to show the great variety and the possible power of these mental attitudes.

The first example will be the famous soliloquy in Hamlet, beginning with the phrase "To be, or not to be." Hamlet is at this point in despair and has come almost to suicide. At the verge he pauses a moment to think. So he is considering killing himself when he says "To be, or not to be." This phrase, when spoken in the first context, will sound quite different to any one whose senses are acute from that spoken in the second context. One easily realizes that each connotes quite a different thing. The feelings of the listener were different in consequence, and so were those of the speaker.

Let us take another example from Shakespere. Shylock says "I stand here on my bond." The situation was that he had just lost his ducats, that his daughter had eloped with a Christian, and that his ships had gone to the bottom. It must also be remembered that a Jew always desires full return on an investment and that he never makes one except upon that expectation. With these things in mind one will understand Shylock's feeling at the moment and see why he was so bitterly eager for revenge upon Antonio. So he turns to the judge and says, "I stand here on my bond." Here one can hardly fail to perceive or imagine the difference between the first and second utterance of the same phrase, and also between this phrase and the line from Hamlet's soliloquy. Now so great a change could not be caused except by a great increase in mental intensity, by a marked alteration in mental attitude. Obviously the change is not in the words, nor is the essence of it in the altered vocal mechanism (for there is undoubtedly a change there), but it is in the mental concept or picture, in the state of mind, in the feeling which animates the words of the speaker and gives them force and power. Furthermore, the listeners are aware of the change, and that fact is utilized by the speaker.

Let us now consider a case where the situation is reversed.

As I sat at the table the other day I asked my wife, who is also a physician, a very complicated physiological question. She hesitated before answering, then proffered this shrewd reply, "That all depends upon the context." She had hardly time to take another breath when my little three-year-old piped up and said "context." She had caught the closing word and repeated it as she heard it. It is surely fair to suppose that the child in using the word for the first time could scarcely have the slightest idea of the concept that goes with it, of what is meant by the word. How completely then it lost its power and how surprisingly it fell short of any intellectual mark.

This last illustration seems to me the best I could give to make my point absolutely clear. The reason is that we have a word uttered by a mind which is able to clothe the word with meaning and to arouse therefore an idea in another mind. The word is then repeated by a third person in such a way that all meaning leaves it and it becomes a mere sound echoed for the pleasure that the mere sound gives. It shows how blank and empty of concept a word can be, how devoid of weight, thought, or implication. These three examples should make clear how the same words can be so uttered as to give totally different impressions, and how these impressions depend upon the attitudes which mold the mind of the debater. Let us turn therefore to a consideration of the principles that govern these mental attitudes.

THE PRINCIPLE OF MENTAL ATTITUDES

Mental attitudes, or states of mind, take numerous forms, and may be divided accordingly. The usual division is that of thought, feeling, and will, or, more strictly in accord with psychological usage, intellect, affection, and will. In considering mental attitudes under these three headings, however, I do not mean to imply that here we have a cut-and-dried, sharply defined boundary between any two. As in psychology, the division is rather a matter of convenience than of absolute separation, and though they are considered apart, two or even three may occur in any single speech, with one dominating the others. Which dominates at a given time depends wholly upon the subject, the effect one has in mind, and the countermoves of his opponent. For convenience we will consider them separately, and at first, at least, in a general way.

1. *Intellect.* Intellectual attitudes take account of many mental processes. They have chiefly to do, however, with statements of fact, with thinking in the ordinary sense, with con-

sideration and reflection, and with innumerable combinations of these same things.

In the mention of bare facts there is merely the mental attitude of unvarnished repetition. This is the simplest and easiest understood of all the mental attitudes. A statement of fact has little behind it except the speaker's determination to make a certain fact absolutely clear to his hearer. The weight and the impression which it gives is simply the weight of its own truth, and the truth has only the effect of inescapable fact itself. To have any effect, facts must be clearly stated and perfectly understood, but beyond clearness they need ask no more. If, for instance, I say "The railroad curves around at this point in a very sharp, abrupt turn and then continues at right angles," the statement is one of pure thought. It is the maintaining of a truth, the facts of which are undisputed, and which has the weight of truth and no more.

Such a statement of fact is manifestly very simple. *Thought* in general is far more complex. Instead of being confined to a single statement it continues to dominate over a series of them, and it goes from a bare statement of fact such as I have already given to the speaker's interpretation of those facts. This interpretation is really the statement of some conclusion to which the speaker has been led by a consideration of the facts themselves. It is affected therefore by the speaker's attitude toward the facts, and it receives credence in an audience in accordance with the degree of influence which the speaker possesses in that audience. If I am President Eliot, my thought probably carries weight; if I am the Kaiser my authority carries much weight, but my thought carries little. As I am, my thought on some subjects has weight, in others none. Thus in an audience the weight of one's opinion is a variable quantity that cannot always be depended upon to make a point. For example, if I say to an audience "It is my idea that the Allies will win," and if the hearers know that I have had no military training, my opinion falls to the ground. Suppose, on the other hand, I say to an audience, "I think his speech defect can be cured in three months." If I have had a better training and a wider experience in speech defects than any other man in this country, my opinion would be received with respect. This is enough to show that a mental attitude may depend for its effectiveness on circumstances outside the speaker's control and that to give expression to it may be dangerous.

One other kind of thought may very well be employed in debate. That is *logical thought*. Logical thinking is not mere

opinion, mere assertion of one's own ideas, but a careful advance by reasoning from one conclusion to another. The audience is led along the same path of thought and brought irresistibly to the same point that the speaker has reached. This is one of our regular debating devices, of course, and needs no further development here.

Consideration is very like thought in nearly all respects. Yet it differs enough for me to call attention to it. As a mental attitude consideration is the same as "mulling a thing over in one's mind." The Bible says, "Consider the lilies of the field, they toil not, neither do they spin, and yet Solomon in all his glory was not arrayed like one of these." Now a remark of that sort is thrown out to be mulled over, as food for thought, so to say. Ultimately, it will be worked up into a conclusion which will move the audience from a previously held opinion. But that final use does not yet come into one's consideration. We are now merely *considering;* as it were digesting the mental food. As a mental attitude consideration is of tremendous value to the speaker. It elicits the attention of his audience, turns its attention to new things, and imperceptibly moves them from a previous position by the simple request to look at a new one. Combined with the point to which it leads and the final step which it induces, consideration is much like the ordinary advertisement. It has been found by psychologists that the ideal advertisement, the one that attracts a reader, holds his attention, and finally leads him to buy, always begins by telling him to consider something of common interest.

So far as the intellectual process is concerned, *reflection* is much like consideration. More exactly, it is consideration for a second time, a looking back, a re-view, a re-consideration. When you assume a reflective attitude of mind and ask your audience to reconsider data already examined, you appeal to their sense of fairness, and you get a cordial hearing because you appeal to this sense. So if you say to them, "Gentlemen, I beg you to reconsider these facts, in a little different light," you secure a new hearing and a new chance for victory.

2. *Affection.* Under the heading of affection we may consider such things as pleasant and upleasant feelings, the being drawn to a matter under consideration or the being repulsed from it. We have there favorable and unfavorable affections. Shylock's desire for revenge, one's love of his country, and the whole range of personal likes and dislikes illustrate the feelings that we have for people and things. May we ask at this point, what of it? Why consider feeling at all? Why do we have the mental attitudes of like and dislike to-

wards any subject-matter? Why not be intellectual and matter-of-fact individuals, devoting our lives merely to enumeration? The debater's reason is this: He wants to make our audience like and dislike. The thing that he wants the audience to like is his side of the question; in fact, he wants so completely to charm them with his interpretation and his presentation that they will like it to the complete exclusion of the other side.

On the other hand, we want that same audience to dislike the opposite side of the question, and we want them to dislike it to such an extent that nobody can ever say anything that will make them like it again. If I am on the affirmative, I am going to try to make my audience like the affirmative and dislike the negative; if I am on the negative, I am going to make my audience charmed with the negative and disgusted with the affirmative. This I hope to do through maintaining an undercurrent of feeling towards my subject, assured that my audience follows through imitation the feelings of the speaker. It may be important at this juncture to explain briefly the reason for this. It is a large subject, and deserves fuller treatment, but I think that a short illustration or two, or, better, reference to the illustrations that have been given before will lead the reader to agree with me that the audience tends to follow the feeling of the speaker.

The first time that I recited Yankee Doodle, the hearer probably had absolutely no feeling in the matter whatever, but the second time, after I had filled my mind with the conception of the funeral and the sadness caused by the death of the boy's mother, he surely did catch the feeling; and notice, he caught that feeling before I had given him any interpretation of the reason why that feeling should be there. That is, he felt the sadness of it without knowing why I had a reason to be sad. This illustration will I believe prove that the audience tends to perceive and reproduce the feeling of the speaker.

Let me pursue the subject a little further. Take for example the experiences through which Shylock passed before the trial. They give cause for the growth of the Jew's intense desire for revenge upon Antonio. Now it is at this point profitable to ask, what is the effect of that intense desire for revenge upon the audience before which he is speaking? It may be one of two. The audience may keep pace with Shylock and completely sympathize with him, as they probably will if he has carried them along naturally to his final climax; or it may do entirely the opposite; it may take sides with the

opposition to Shylock, and agree entirely with those who plead for mercy. However, as I have watched audiences at this play, feeling has always wavered; that is, the audience sometimes has followed Shylock and sometimes followed the opposition, usually adhering to Shylock if he is not too severe, and then going over to the opposition if it is not too weak in its appeal for mercy. All this makes "The Merchant of Venice" an extraordinary example of the very thing I am trying to illustrate, for it shows how an audience tends to follow the feelings of the speaker, inclining toward revenge or mercy as either is forced upon them by the actor.

Let this suffice for our consideration of the mental attitude of feeling. It is a difficult thing to attempt before an audience, for so to affect people that they will feel kindly toward one's own side of the debate and unfriendly toward the opposite is a subtle and delicate mental performance. Yet difficult as it is, it is worth while. Usually it means victory if one can so accomplish it at the outset that throughout the rest of the whole debate one's hearers will maintain a liking for one's own side and a dislike for the other. I know that oratory is commonly scorned by the academic trainers of speech, but this is not because real oratory deserves scorn. It is because the academician fails to understand the depth to which the non-academic oratorical principles penetrate. I am fitted to talk upon this subject, because I have been trained by both schools, by the academic and the non-academic alike. I am not at all, to be sure, a typical product of either, for I am absolutely out of practice. I can, however, as a teacher bring to others results to which I cannot myself attain. My study in both of these fields leads me to the following understanding of the situation.

The academic school has emphasized throughout in its method of education and training what I call the external method; that is, the word form, the external enunciation, the intellectual cast of the language itself, with program of argument and tabulation of statements and points. On these things they lay the chief emphasis. The non-academic school, on the other hand, proceeds from within outward. They make their chief aim the development of thought processes, the development of power to visualize and picture ideas and conceptions and to throw them clearly upon the screen of the mind. To do this they must of course have a means of expression; they must have speech to tell it all, and words to describe. This they do have, but by no means do they put their chief emphasis on words. Be it far from me to condemn

either of these schools. They both have their value and charm; and they both have these to such an extent that the whole matter of speech in debate is seriously wanting where it falls short of possessing all the excellences of both.

I am also aware of the fact that each of these schools criticises the other unfavorably. The academic school, or the school of the external word, criticises the other because it lacks good English, lacks expressiveness in words, lacks power to put in the word itself all shades of meaning; because it demonstrates too much feeling in its utterances, ("gush," as they call it) and so on. The non-academic school, or the school of the internal thought, criticises the academic because it puts too much reliance upon the ability of words to carry ideas; because it uses no other methods of external expression than language; because it fails to utter the same series of words now in a simple fashion, now with the atmosphere of melancholy interpretation, and now with the blue sky of a sunny disposition, changing the view completely. The non-academic school also finds fault because the other fails to give in its renderings much else besides the word; because it fails to give shades of thought; because it fails to give minute coloring of interpretation; because it fails to give an overshadowing feeling, a tremendous madness, or the utterance of an insatiable revenge, no matter what the words may be. My study in both of these schools has taught me that the non-academic school has by far the best of the argument. I grant that the non-academic school lacks in vocabulary, fails in the presentation of completely polished English, and may sometimes be guilty of the seemingly unpardonable rhetorical sin of using words not strictly according to the dictionary rules; but I maintain that the excellences of the non-academic school can outnumber and outweigh all these little, minor, and pardonable faults.

Some of the things that the non-academic school can do which the academic school cannot do are these: The non-academic school can maintain an extreme variety of mental attitudes over and above any words that one can give them to say. Those mental attitudes they can vary according to the state of mind into which they choose to put their audience, with all the shades of feeling, intellect, and will. They have a tremendous power and sweep of the imagination, and they have their creative imagination so developed that they can picture vast conceptions on the spot. And they can do not only these things; beside all this and to a degree that I have never seen in any academic debater, they can read their audience

and judge from its action what sort of an effect they are making, and from this reading shift their arguments, and move their debating procedures accordingly. Not only can they read their audiences, but they can comprehend the mental attitudes in the speakers of the opposition to a remarkable degree. In fact, they can look through and above the words the man utters and understand his intellectual coloring, understand his imaginative interpretations, understand his shade of feeling in the way that no purely academically trained debater can possibly do.

In brief, the academic people fail to appreciate the introduction of feeling into debates. They believe that usually debate should be without feeling, and, as far as possible, entirely unemotional. The other school insists that emotion has its place, its service, and is one of the most valuable elements that can be introduced in debate if properly and opportunely employed. From all this I may draw the conclusion that the mental attitude of emotion has an unusual and subtle power in debate and deserves serious consideration and cultivation.

3. *Will.* Will is psychologically choice, and the employment of will as a mental attitude means an appeal to the will or an appeal to the choice of a hearer. You know that you can persuade him; you know that sometimes you can drive him; you know that sometimes through feeling you can so elicit his sympathy that he will be led entirely by it. But just here we have a definite requirement. You must lead him to his own choice. That is, you must lead him to what he thinks is his own choice, but really what you, by your skilful presentation, have led him to choose. He must not know of your leadership; he must only know that you present facts, interpretations, likes and dislikes in such a way that he seems himself to choose; and of course you wish your hearer to choose the side that is favorable to yourself.

To lead another to choose is a very complex and difficult matter, but it is an object that one must try to attain all the time one is arguing. If the end and aim of one's speech is to appeal to the hearer to make his choice one must make clear statements of fact, must persuade as far as one can to a point of decision, must reform the hearer's consideration of one's subject, must guide and prejudice his reflection in such a way that he cannot be drawn to later interpretations of an opposite nature, one must appeal to his feeling of patriotism, to his love of home, to his honor, his sense of justice, and his hatred of every thing that is wrong, in order to force his decision and force it to one's side. Then, lastly, the speaker by his mental

attitude of will must force his hearer to make a final change from his own preliminary choice in the matter, to the interpretation presented by the speaker.

To sum up, the mere words of the speaker are the smallest part of the ideas, concepts, and feelings, which his mind seeks to present. Behind the words he possesses knowledge, makes reflections, maintains mental states, develops mental attitudes, and expresses all of these more or less independently of the words, and reproduces them, as far as the hearer is sensitive enough to perceive, in their original form within the minds of the audience.

A CASE OF INTELLECTUAL DEVELOPMENT DESPITE ENFORCED SECLUSION

By Josephine Curtis Foster, Chief Psychologist, Psychopathic Hospital, Boston

The psychologist who gives mental tests is frequently met by the statement "This low mental age means nothing because the child has been sick and has not been a regular attendant at school" or by "You can't expect a child who has lived on a lonely farm and who has had no playmates to test at age." In vain may the psychologist answer "I admit that inability to read may lower the mental age with a scale like the Binet or the Point Scale. In other respects, however, school training has little influence on the outcome of psychological examinations." Such a reply usually fails to carry conviction to one who is not an experienced examiner. Instead, an objector is likely to depart, thinking that the psychologist is grossly exaggerating the importance of the tests, resolving to take mental ages in the future with as large a grain of salt as ever, and continuing to consider the mental age only when it agrees with his personal opinion of the case.

In our work at this hospital we have come to use the answer "Let me tell you about a fifteen year old boy we examined some months ago." The illustration has been so successful that we offer it here to other psychologists for similar use.

The Story of John

John Henry [1] was born in a rather poor district of a New England city on August 13, 1902. Of his remote ancestors we know comparatively little. Presumably his paternal grandparents were natives of Ireland. His maternal grandparents are said by one of our informants to have come from England and Scotland, and by another to have been born in Nova Scotia and Maine. We have one history which says that a maternal uncle died in an insane asylum. Another maternal uncle, who came to the hospital upon request, is reported to have been insane, appeared to be very unintelligent and was considered

[1] This, of course, is not the real name, but it is like it in having the same second name (which the boy spelled Henery).

as "probably untrustworthy in his statements." This uncle denied mental illnesses in the family, but at the same time showed such familiarity with hospitals for the insane as to suggest personal acquaintance. Of the immediate family, the boy's father was born in Ireland, and was probably a horse-trainer. He was heavily alcoholic, abused the children, and some five years ago deserted the family. John's mother was born in New England. She was at this hospital for a time, and is now an inmate of a state hospital for the insane with a diagnosis of Dementia Praecox, paranoid. Four of John's siblings are said to have died of starvation in the first year. An older half-sister is apparently normal.

The schooling of John and this half-sister, whom we shall call Julia, was decidedly meagre. They were allowed to attend occasionally, but although both of them learned to read, they acquired little else. If we could add together all the days John spent in school, the total might come to three years. We know that the family moved so frequently (at least eleven times in four years) that he could seldom have spent a full term in any one school. John was last in the fourth grade. He attended school in this grade irregularly for some years. Apparently he was not held from advancement on account of dulness, but because the mother went to the teacher and said "Don't waste any time over this boy of mine. He is feeble-minded and will never learn, no matter how much effort you expend on him." Apparently, also, the teacher believed the mother who was so candid about the failings of her son, accepted her advice, and spent her energies on the other children.

As might be expected under the guidance of such a mother, home life was not an aid to John's mental development. Julia says that her earliest remembrances are of squalid surroundings, a room or two with the poorest of furnishings, which were never anything but filthy. She had very little clothing, and at times such a small amount that she was forced to stay indoors. John says that his mother would buy him new clothes and then lock them up till he had outgrown them. Moreover, there was never enough food, and the children were continually hungry. The mother and step-father beat Julia and obliged her to chop wood and do all the work. In her presence they also indulged in perverse sexual practices. John was not abused to the same extent, probably because he was the father's own child. However, the only attempt to teach him anything was his father's instruction in obscene verses and songs.

John and Julia had not even the opportunity to learn from

contact with other children. Their mother often kept them in the house, and this happened still more frequently after the father deserted. Even when the mother was away scrubbing in the middle of the day, the children dared not venture out, for fear there might be some truth in their mother's statements that "there was a God lover head" who would do various things to them, that the policemen would get them and do terrible things, that people would kill them, and later, when Julia was living elsewhere, the mother told John that "they would catch him in the draft"—a hazy, but horrible fate.

John's story of his life is of a succession of moves from house after house, with periods of confinement varying from a few days to one that lasted 13 months. Most of the daytime he slept or amused himself by reading the Bible, the few religious books and the few Indian stories which they owned. When the mother returned from her work, he seems to have been occupied with endeavoring to escape her wrath. He says that she would ask him questions and then answer them herself. If he attempted a reply, she usually slapped his face. She talked continually of the references which the newspapers made to her, and of the pictures the moving-picture men were taking of her. They had no regular meals. The mother probably ate something at the restaurant where she did scrubbing, but all she brought home to John was bologna sausage, milk, crackers, and bread. When he came to this hospital, the boy had not had a hot meal in months. At night he was often kept awake by his mother's talking, by what he called "Mother's lies."

Things went from bad to worse with John, but they improved for Julia in the spring of 1917. At that time the mother accepted a position as "an experienced cook" and took the children with her. On the third day she was discharged as dirty and hopelessly incompetent. Julia, however, was retained and since that time has had normal surroundings. John and his mother returned to their squalid living, and the boy was confined even more closely until early in 1918, when the police and one of the children's societies found them and brought them both to this hospital.

When John was examined here, he was a large, overgrown boy, sallow and anemic. His voice was husky and uncertain. He did not know how to meet strangers, and was shy and reserved. His face was a little sober, but almost expressionless. He smiled slightly and then only after many interviews. He answered all questions as briefly as possible. The most surprising thing about his attitude was that he took practically

no interest in his surroundings or in what was to happen to him. John's ignorance of the world about him was astounding. He could name the president, the governor, and several recent mayors, but when asked for the capital of Massachusetts he did not understand what was meant. He knew nothing whatever of geography. With the exception of reading, arithmetic was the only school subject in which his performance was at all creditable. He had never seen a moving picture or a baseball game.

We have here, then, a boy of fifteen and a half, who has had practically no school training, practically no contact with other children, and no home education save what he may have picked up from the reading of a few religious books, and perhaps a stray copy of a newspaper.

In spite of this overwhelming lack of information and of training, when given a psychological examination John received a mental age of 16.0 (I. Q., 1.04) on the Yerkes-Bridges Point Scale, and a mental age of 15 yr. 1 mo. (I. Q., .98) on the Stanford Revision of the Binet. There is nothing in the examinations that would be unusual in a normal boy of his age. On the Point Scale the errors he made were in: repetition of 7 digits, repetition of the longest sentence, failure to give interpretation for two of the pictures, failure to give definitions in terms of classification, acceptance of the line suggestion, failure in one of the memory drawings, in one of the definitions of abstract terms, and in one analogy. On the Stanford scale, he passed all the 9 year tests, failed in designs from memory in year 10, in interpretation of pictures in year 12, in the vocabulary in year 14 and the succeeding years, in giving differences between abstract words, in repeating backwards, and in the code in year 16, and in the repetition of digits forwards and backwards and in giving the thought of a passage in year 18. His arithmetical ability is particularly evident in the Stanford examination. He did the Healy Construction Puzzle A in 7 moves in 15 seconds and Construction Puzzle B in 12 moves in 30 seconds. His attitude during the psychological examination was somewhat listless and a bit childish, but he cooperated well and showed occasional interest and pleasure.

The conclusion must be that, in this case, at least, lack of school-education, lack of home-training, and lack of normal intercourse with other children have not prevented a normal rating on intelligence scales.

John's subsequent history, while not of consequence for our purpose, is, nevertheless, interesting. He was discharged from this hospital

to a society for placing and caring for children, with the diagnosis "Not Insane, not Feeble-minded." We hoped that it would be possible to give an examination at the expiration of his first year of living in normal surroundings. This proved inadvisable because the boy is doing good work on a farm, and a trip to the city might encourage his present ambition to work in a store, a job which doubtless would not fit him just now. The society, however, sent us a detailed report of a "visit" made and we give the substance of that report here:

John is living on a farm with a family who understand his case and are anxious to assist him in every way possible. He has gained in weight since last year and he looks physically well. However, he is very slow in his movements, and appears to take no interest in his surroundings. He seemed little affected by the fact that the visitor had come from Boston to see him, and the family say that this is his usual attitude. In the year that he has been with them, he has never spoken of his mother and sister, has never asked questions about places to which he has been taken in the automobile, or shown any interest in them. The only exception was one trip to the ocean, when he asked "Why can't you see the other side?" But he has never spoken of the trip since. His typical day is as follows: Rises at 6, without being called, feeds the horses, without being told, milks three cows, comes into the house for breakfast, without being called, eats a hearty breakfast, then without any comment, returns to work, waters the cows, and cleans up the barn. He comes back to the house about meal time, gets the mail from a box at the end of the lane without reading any addresses or ever asking if there is a letter for him. After the noon meal he takes care of the hens, and does odd jobs. In the summer time he goes for the cows. After supper he reads the jokes in the daily paper and Farm Journal, sometimes smiling a little. He never grumbles while doing anything he is asked, never asks a question, very seldom speaks without being spoken to, goes to bed without being told to at 6:30 and is perfectly willing to do this day in and day out. In fact, far from disliking routine, he is somewhat disturbed when rain prevents the performance of some of his usual chores. He shows no interest in anything. He is good to the animals but never pets them and the dog and cat never ask him for food. He never notices anything unusual. If, for example, a cow were in great distress and he had not been told to notice whether or not the cow was sick, he would make no comment. Neither is he at all interested in the fact that the farmer is willing to give him $50 if he remains with him a full year. Besides this total indifference, the farmer and his wife have reported several facts which are of importance in the prognosis of the case. He has recently been smiling more often, and occasionally laughs heartily when alone at his work. When asked what he was amused at, he insists that he did not laugh. He has never made any attempt to play with a boy who lives in the next house, a quarter of a mile away. It seems likely that the boy may end by having Dementia Praecox, like his mother.

BRIEFER STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF McLEAN HOSPITAL

Reported by F. L. WELLS

		PAGE
I.	The "Coachproof" Test	172
II.	A Filing Test	178
III.	A Brief Test for Mental Accuracy	179
IV.	The "Cued Combination" Method	183
V.	Geographical Information in High School and College Students	186
VI.	An "Omnibus" Test for Reasoning and Imagery	189
VII.	Certain Individual Records Compared with Standard File	191

(The experimental work of these studies was complete on July 15, 1917.)

I. THE "COACHPROOF" TEST

1. The "clinic rounder," vitiating Binet findings by previous experience of the tests, early demanded the protection of such measurements against foreknowledge. As mental examinations are introduced in various kinds of personnel selection, and it becomes increasingly to one's interest to make a good score, protection against coaching in the methods used becomes altogether necessary.

2. A test is "coachable" to the extent that superior specific ability in it may be acquired through practise in it, or through knowledge of the result to be attained. Tests of involuntary processes are least subject to coaching. Sensory tests are somewhat coachable, though it is questioned whether the improvement is in strictly sensory acuity, or in more efficient use of accessory cues. Voluntary motor tests are coachable to the extent of practise improvability, which may be considerable if not rapid. With well designed experiments, the practical danger of coaching in these fields is not great. The danger of coaching is greater in the higher mental processes, where right response is determined by factors of conscious knowledge. One simply learns that he is to speak certain words or draw certain lines, in response to particular problems. While such coaching may be unmasked in individual examination, it is none the less vitiating thereto. In group experiments it is less likely of detection.

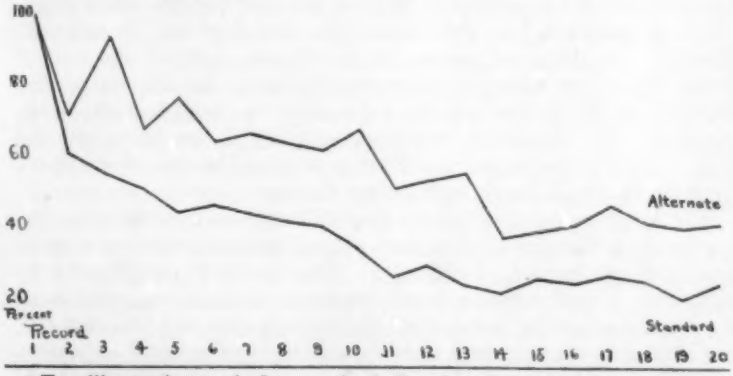
3. There is little protection against special ability acquired by practise in the actual motor or sensory tests. Protection against coaching in tests of the higher mental processes is provided through alternative tests of similar material, but so differing that the literal responses to one are not correct to

another. Tests with a number of homogeneous alternatives are for practical purposes regarded as commensurable and coachproof.

4. Though essential test elements are changed in the alternatives, they preserve common factors open to mutual practise effect. Such tests are still coachable to the extent to which practise in any one form of the test will improve performance in its alternatives. In this way is measured the degree to which a test is coachproof. This problem is essentially the not unfamiliar one of practise and transfer.

5. In the course of preparing material for coachproof instruments of mental examination, inquiry was made into the perfection with which the method of alternation attained this purpose. Ten alternative forms of the Woodworth-Wells hard directions test were prepared.[1] A subject accomplished one of these and one of the Woodworth-Wells standard blanks daily for ten days. On odd days the standard form preceded, on even days the alternative form. Then ten standard blanks and ten alternative forms were on the same day accomplished in immediate succession. There was thus practise in an identical task, paralleled by practise in a continually changing task. Evidence of coachability in the test is the appearance of practise in performance with the alternatives.

6. The score is in seconds' time. For each error, the time consumed is increased by 1/15 of itself (there are 15 responses in the test.) With this correction for error, the initial time of the standard form is 92 seconds, of the alternative form 94 seconds. Figuring these starting points as 100 per cent, the per cents of each reached in the subsequent tests are shown in the accompanying curves.



[1] For illustrations, cf. *Journ. Appl. Psychol.*, 1917, vol. 1, p. 139.

7. Both standard and alternative tests show significant practise improvement at the end of twenty performances. The best performance in the standard test is 24 per cent of its original score. The best performance in the alternative test is 40 per cent of its original score. The subject's performance at the beginning is about the normal average. After practise periods totalling less than an hour the alternative forms are accomplished in time near the lower limit of the unpractised normal range. A test having this property is not well fitted for use where there is special motive to make a good showing, and normal opportunity to train therefor.

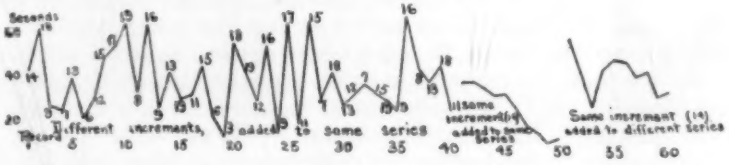
8. Other experiments were made in which the practise was not obtained from day to day but in immediately successive performances. Practise was given in a "constant increment" test, adding 17 to each of a series of 10 two place numbers. Time and errors of each series were noted, the experiment consisting of 45 series. The first three series were different, and average 36 seconds without correction for one error made. The third series was then practised for seven records, without error, the time decreasing to 21 seconds. The subject having then studied for 3 minutes to learn the series, practise was continued to the fifteenth record without change. Then new series (Nos. 4-8) were introduced for five records. They were accomplished without error, in average time of 29 seconds. Then series Nos. 1-8 were repeated (records 21-28). Series 1 and 2 showed about 68 per cent of their beginning rate, the others were longer than their previous records. Record 29, a wholly new series (No. 9) is 28 seconds and errorless, clearly better than initial performance. Practise being resumed on series No. 3, (records 30-35) the subject was requested to memorize this series and report when done, (224 seconds). The errorless time dropped to 14 seconds. Records 36, 37, new series, have 68 per cent of the initial score; 38 is very long, apparently blocked. In the remaining records, 39-45, series 1-7 are used with the addition of 16 instead of 17. Practise improvement begins to show at the close of the experiment. There was subjective feeling of "ready to stop," as though doing slowly.

9. Continuous experiments like this are objectionable as introducing a fatigue (refractory phase) element not present in practical application of the tests. This tends to mask gains by transfer. On the other hand, there is an accustomedness to the whole physical setting of the experiment that is helpful, but absent in the conditions of practical application. A somewhat more satisfactory situation is indicated than in the hard directions test. The best performance with the same incre-

ment in a new series is 74 per cent of initial performance. The best performance in the practised series is 41 per cent of initial performance, representing the time of writing a memorized series. When the increment is varied (adding 16 instead of 17), the gain is practically neutralized.

10. Another subject accomplished 40 records in a constant increment test from a single series, No. 4. In these records different increments were added to the series. These numbers added ranged from 3 to 19. The increment in the initial series was 14, and during these 40 records this number is not so employed again. In records 41-50, the increment 14 was added to series No. 4 in all cases, thus giving practise on this series and number only. In records 51-60, 14 was added to different series, save that the last two records were again with series No. 4.

11. The score is the time consumed for a series, plus 10 per cent thereof for each error. These scores for the 60 records are shown in the accompanying graph. Attaching to each point of the curve for the first 40 records is a special figure. This denotes the figure added in that record. In records 41-60, 14 is always added; to the same series in records 41-50, to different series in records 51-60.



12. The practise in adding other figures to series No. 4 does not appear to increase facility in adding 14 to series No. 4. In records 41-50, the repeated adding of 14 to series No. 4 rapidly improves capacity in that specific function. In records 51-60 it is not evident that the previous 50 records have improved the adding of 14 to any other series than the one used in those 50 records. It results that practise with the same series and the same increment gives great improvement. Different series with the same increment give some improvement. The same series with different increments gives very little if any improvement. In practical experiments it is easy to vary both series and increment. It is indicated that superior capacity for an unforeseen series and increment is not to be gained save through a generally increased capacity for all series and all increments. A test with this property is coachproof.

13. Other experiments were performed in which the task is the rearrangement of various items in a serial order. A

subject was instructed to rewrite in order of magnitude a series of ten 4 and 5 place numbers. The first two records were with different series. Series No. 2 was then practised until the tenth record. The 11th record was a new series. Practise was continued with Series No. 2 until the twenty-first record. Nine new series were then introduced, making thirty records in all.

14. The score is the time in seconds for accomplishing a series, plus 20 per cent of itself for each break in the series, plus 10 per cent of itself for each lapse in writing. Fourteen out of the thirty records taken were penalized for error.

15. Records 1 and 2 are errorless and average 46 seconds. The last two errorless practise records, (19 and 21) average 46 per cent of this score. The last two errorless new series, (records 25, 26), average 40 seconds, 87 per cent of the initial score. The final group of new series, (records 22-30) averages 130 per cent of the original score, so heavily were they penalized for errors. No marked gain in the accomplishment of new material is indicated.

16. In a subsequent experiment the procedure was changed in that the subject instead of writing the series itself in order, numbered the items to show the proper order.

17. The score is the time consumed in accomplishing a series, plus 20 per cent for each break in the series. Of the 25 series accomplished, 14 are without error.

18. The initial time is 35 seconds, the score being raised by one error to 42. The responses improve quickly through memorization. The seventh record is 19 per cent of the original score, beyond which no improvement is observed. The best score in a new series is 52 per cent of the original score, (record 17). The best score in a new series penalized for error is 73 per cent of the original (record 21). It is indicated that the practise effects some general improvement; less coachability appears than in the hard directions test and more than in the constant increment test.

19. In another subject, the experiment was performed not with numbers, but with names to be numbered in order of alphabetical sequence. The first two records were with different series. Series No. 2 was then practised until the ninth record. The tenth record was a new series. Practise with series No. 2 was then continued until and including the twentieth record. Ten new series were then introduced, making thirty series in all.

20. The score is the time in seconds for accomplishing a series, plus 20 per cent of itself for each error. Six of the thirty series were thus penalized.

22. The improvement in the practise series is to 15 per cent of the original score. The best score in a new series is 55 per cent of the original score, (record 30). The best score in a series penalized for error is 65 per cent of original score (record 22). Similar coachability to that with the consecutive numbers is indicated.

23. Attention is invited to already reported results on practise with the free association experiment.[2] With no repetition of stimulus words there are general decrease in time, and other improvements of performance. Records of this test may be vitiated where there is motive to attain high degree of practise. There are also certain instructions about responding whose observance would improve the score if not unmasked in lengthened times and the character of responses.

24. The number-checking test[3] employed in connection with the above association experiments is not alternative and thus without significance for the present question. The addition material (Kræpelin's *Rechenhefte*) is alternative and shows considerable practise. It is not probable that the subjects' general quickness at figures is increased so greatly if at all. For quickness at figures, such a test is invalidated in special practise, the increased capacity being confined to a particular and practically unusual sort of adding.

25. Such a test of adding is very analytical; it does not cover the field of quickness at figures, but a specific part of it which will normally be an index of the whole. If a subject is specially practised, this relation is distorted. As an analytic test, a test is "coachable" in any form in which it shows practise improvement.

26. Alternative "typewriting tests" are coachproof in spite of their practise improvement, because such improvement applies more equally over the general accomplishment of typewriting. Instead of analysis, the typewriting test aims at synthesis, that is, at reproducing in miniature as much as is possible of the practical accomplishment. Where, as in typewriting, a reasonably synthetic test can be made, "new" test material is also coachproof.

27. The practical significance of mental tests, except the specifically "trade" tests, is analytical. It is indicated that alternation leaves certain tests open to not inconsiderable practise improvement. Such tests can still be "beaten" where the practical motive is sufficiently strong.

F. L. WELLS.

---

[2] Practice Effects in Free Association, *Am Jour. Psych.*, 1911, vol. 22, 1-13.

[3] Relation of Practice to Individual Differences, *Am. Jour. Psych.*, 1912, vol. 23, 75-88.

II. A FILING TEST

28. The essential operations of filing are searching the files for a given entry, and filing a new or removed entry in the file. Efficiency of file clerks is given in the speed and accuracy with which these operations are accomplished.

29. The part played by memory, patience, and general intelligence in this accomplishment, increases with the complexity of the files to be handled. The filing test proper is concerned with specific aptitude in the psychomotor processes of manipulating the file. The task is to find the place where a given entry is or ought to be.

30. A testing file of 500 cards, 2½ by 3 inches, contained in a standard file drawer, was furnished for the present purpose by the Phœnix Mutual Life Insurance Company.

31. For the test of searching fifteen cards at random were removed from the file, and five other names in it were noted. All twenty names were written on a single sheet, which was handed to the subject, with instructions to find whether the names were present in the file or not. If the name was found, this was announced by speaking the name; if not, the announcement of the name was preceded by the word "No." Another procedure was to have the names writen one each on memorandum slips and have the subject verify them as above. This made it possible to present the same names in varying order. Each operation was timed with a stopwatch.

32. For the test of "filing" proper, twenty cards at random were removed from the file. They were given to the subject with the direction "File these cards in their right positions. Let them rest simply on the rod" (the locking rod as shown) "without removing the rod and pushing them down into place." The time of filing each successive name was noted. Afterwards the subject might be asked to check the correctness of the cards just filed, a process taking from 50 to 170 seconds. The correctness of the filing was verified before removing the locking rod and returning the cards to their final positions.

33. Results from eight subjects are as follows:

Average time in Seconds in Elementary Filing Operations.

Subject	A	B	C	D	E	F	G	H	Av.	ratio of M. V.
Searching	15	21	25	19	35	25	19	18	22.1	.21
Filing	25	17	14	19	37	26	17	15	21.2	.28

34. The first five subjects are without special practise in office methods; the last three are office secretaries, but not professional file clerks. There is evidence that such times as

the above are much reduced in special training. There is general correspondence in the speeds of searching and filing, which is confirmed in office experience. Subjects A and C however, are quick in one operation and slow in the other.

35. Final scores are stated in terms of time. Errors do not essentially affect the present results. In general experimental practise, it is best to correct a time score for errors through increasing it by a percentage of itself. Heavy penalty is to be imposed for errors in a filing test, such as ten per cent of the time score for each per cent of error.

36. File manipulation is influenced by external factors such as the number of names in the file and their dispersion through the alphabet. The names in the present file included but a small fraction of the letter A. A file with the same number of names but running throughout the alphabet, would probably be easier of manipulation.

37. The above represents a standardizable form of test for filing operations proper. It is coachproof. It was in the present instance applied as an individual test, but is convertible into a group test with sufficiently extensive testing files.

F. L. WELLS.

III. A BRIEF TEST FOR MENTAL ACCURACY

38. It was desired to arrange series of tests which should be (1) coachproof, (2) suitable for application to groups, (3) applicable to individuals with less demand upon the experimenter's attention than time-limit group tests. This and the following studies deal with tests so arranged.[1] In practical application these tests are most suited for the selection of clerical workers.

1. *Add 17 to each of these numbers. Write each answer on the dotted line to the right of each number; thus* 10 27.		2. *Divide*	3. *Rewrite this list of numbers, putting the largest number on the dotted line at the top, and the others in order of size below it.*	
34.....	31.....	384) 74496	8846	.....
79.....	48.....		9059	.....
88.....	66.....	761) 195577	10604	.....
68.....	26.....		10653	.....
25.....	60.....		10097	.....
82.....	98.....		10578	.....
27.....	37.....		10585	.....
30.....	13.....		8605	.....
24.....	52.....		7895	.....
23.....	78.....		8813	.....

[1] A brief time limit intelligence test was also devised, described by G. G. Fernald. Report of the Maine Commission for the Feebleminded, 1918, pp. 11, 75.

4. *On the dotted line before the name coming first in alphabetical order, write the figure* 1; *before the name next in alphabetical order write the figure* 2; *and so on until you write* 10 *before the name last in alphabetical order.*

..... Foley, Annie E.
..... Gates, Mary M.
..... Kennedy, Allan T.
..... Mac Donald, Alice J.
..... Robertson, Clara L.
..... MacDougall, Vincent
..... Judge, Mary
..... Mingo, Ida L.
..... Marden, Edith
..... Roberts, Fannie

5. *Copy the numbers given below, putting them on the dotted lines at the right in the same order as they are given.*

514863729 . . . . .
169472853 . . . . .
384259617 . . . . .
837594162 . . . . .
392617584 . . . . .
471862953 . . . . .

	A	B	C	D	E	F	G	H	J	K	L	M	N	P	Q	R	S	T	U	W
1																				
2																				
3		x																		
4																				
5		c																		
6	n																			
7				g																
8																				
9															t					
10																		f		

6. *Each of the squares in the diagram at the left is named according to the letter under which it stands and the number on the same line with it. Thus the square with an* x *in it is named* B3.

Give the letter and number to name the squares which have in them the small letters c....; f....; t....; g....; n.....

Put a figure 6 *in square* W2
" " 8 " " N4
" " 2 " " F3
" " 7 " " G4
" " 4 " " J7

7. *Do what it says to do and be sure you get just what it says.*

If New Year's comes in May, make a figure 6 here....; but if not tell where the sun rises..... Give a wrong answer to this question, 'How many days are there in September?'.... Write "no" no matter whether Canada is in Africa or not..... Write the second letter of your first name and the first letter of your last name at the beginning of the dotted line..... Notice the numbers 2, 7—If iron is heavier than water, write the smaller number here....; but if iron is lighter, write the larger number here.... If people believe that Columbus was King of England, cross out what you last wrote, but if not, put in the number to complete this sentence, 'A cat has .... ears.' If Friday comes after Thursday, make two crosses here.....; but if not, make a square here.....or a circle here..... Give the correct answer, "yes" or "no," to this question, 'Do apples have four corners?'..... Write again what you last wrote, here..... Make a comma after the shortest of these three words, hat pill ladder. Put a square before this name of a boy.....John. Write any letter except "e" just after this comma, ... Make a dot to the right of any one of these letters F G H I J. Show by a cross when the days are shorter: in winter?.... in summer?.... Write "no" if 2 x 6 are 12.....; if not, make a circle here.....

39. The present study concerns the first of these test-series, termed C. A form of it is reproduced herewith. It is seen to be capable of indefinite alternation. Actually ten forms in

addition to the above have been printed, omitting however test 7, the alternative of the Woodworth-Wells Hard Directions test.

40. It is intended that the problems of this test-series shall all be well within the range of the subject's information and experience. The quality of performance then depends in first instance upon the *care* with which the tests are accomplished, not upon special information of correct answers or ability to reason them out. In subjects for whom the series is suited, performance is limited essentially by the subject's carefulness. Other test-series are limited by knowledge and by reasoning ability.

41. For accuracy the tests are scored as a point scale. The sum of possible credits is 100. Each item represents a proportional part of these credits. Deductions therefrom are made for errors uncorrected. In tests 3 and 4, a deduction is made for corrected errors, which is doubled for uncorrected.

42. A rational final score in tests of this nature, or any function thereof, is the per cent of all records in file which the record in question surpasses. If it is the best of all records, it thus scores 100 per cent; if better than half, only 50 per cent, etc. A card file of results is kept, in which each new accession to the data is entered in order of its amount of credit.

43. The present data concern 140 individuals; 75 records from college students, 53 from high school students,[4] the remainder from adults of the same order of intellectual capacity as these. From the college group, the per cents surpassing different scores in points earned, and in minutes of work, are as follows:

Point Score	Per cent surpassed by same	Time for test (min.)	Per cent surpassed by same
20	0	22	0
30	1	20	1
35	3	19	4
45	4	18	8
50	11	17	13
55	17	16	20
60	24	15	28
65	29	14	38
70	40	13	53
75	52	12	67
80	61	11	82
85	73	10	94
90	84	9	98
95	90	8	99
100	97		

[4] For this material the writers are indebted to the courtesy of Dr. H. T. Woolley.

44. With the high school group, the time score was in terms of points earned per minute of work. In this group, percentile values of various scores are approximately as follows:

Point Score	Per cent surpassed by same	Points earned per minute	Per cent surpassed by same
31	0	1.0	0
40	8	1.5	4
50	15	2.0	4
55	19	2.5	7
60	32	3.0	19
65	40	3.5	32
70	47	4.0	45
75	54	4.5	60
80	60	5.0	68
85	74	5.5	79
90	83	6.0	87
95	93	6.5	91
100	96	7.0	95
		7.5	98

45. When graphs of scores for both groups are superposed, there is evident no separation of the groups in respect to number of points earned. The ranges of performance are similar in extent. The number of points earned per minute is clearly greater in the college group. Their superior mental discipline enables them to perform the operations faster, but they are accomplished with equal accuracy in both groups.

46. The accompanying table analyses the error penalties in different tests of the series.

PER CENT MAKING

	No errors	1 error	2 errors	3 or more errors
1. Addition	48	25	10	17
2. Long Division	73	21	6	..
3. Consecutive numbers	44	21	15	20
4. Alphabetical names	50	14	8	28
5. Copying	93	4	3	..
6. Letter-square	60	19	3	18
7. Directions	27	36	21	16

47. As is consistent with paragraph 45, the different items are of nearly equal difficulty for the high school and college groups.

C. M. KELLEY,
F. L. WELLS.

IV. THE "CUED COMBINATION" METHOD

48. The present study deals with performance of high school and college students as indicated in paragraph 43. The test-series is termed K, the form used being reproduced on the page opposite. It employs a variant already described[5] of the combination method, and is a preliminary form of the test-series, to which similar alternatives have been prepared.

49. The first portion consists of words 35-70 inclusive of the Terman Vocabulary Test, omitting the words *stave, snip,* and *swaddle*. The cues in this portion are intended to be such that the word cannot be supplied unless known. The second portion contains words of special difficulty in spelling. The cues for these words are intended to indicate the word clearly, without indicating its correct spelling. The third portion presents another kind of material, and is discussed subsequently.

50. This test-series differs from series C in that it is limited by the subject's absolute knowledge of correct responses rather than by his carefulness in setting them down. Carelessness is less likely to vitiate what is known here than in series C, because there are more logical connections by which responses check themselves up. Carefulness adequate to perfect score in series C avails nothing for the responses of series K outside the range of knowledge. Practically, this type of test-series is indicated where clerical duties take on a secretarial nature, more intellectual than reached by the mechanical tests of series C.

1. *Each of the sentences below has a word of which only certain letters are given. You are to fill in whatever letters are missing so that the word is complete. If you are not sure you have filled it in rightly, you may put a ? after your answer. If you cannot fill in a word in about 15 seconds, go on to the next.*

Each period (.) *in a word means one letter left out; a long dash* (———) *means any number of letters left out.*

Examples:

The boy was sucking a big ripe or. . . e (orange).
A cent is made of co———(copper).

This policy will .ns. . . your house against fire. We r.g..d your idea very favorably. The dentist was obliged to kill the n. . . . of the tooth. We heard the carriage-wheels c..nch the gravel. The j. . . . .r kept six balls going in the air at once. The judges in their robes represented the m. . .sty of the law. She is of a deep brun. . . . complexion. He copied his superiors in an a..sh manner. The pup is a sp..ti.., frolicking little animal. She laughed herself into h. . . . . cs. M. . . was

[5] *Journ. Appl. Psychol.*, 1917, Vol. 1, p. 140. *Journ. Educ. Psychol.*, 1917, Vol. 8, pp. 483-487.

the Roman war-god and is the name of a star. He left to seek r.p..e from his labor. You are too shr... to be tricked by such talk. Who loses the game must pay the f..f..t. He has a marked p.c.....ty of speech. The c....ge of 2½ cent pieces has been thought of by the government. The table was inlaid with beautiful mos... designs. They ceased to bew... their misfortunes. The labor of doing that is disp.....nate to the good it does. The house, old and d...p.....d, is still an interesting ruin. The company was granted a ch....r by the state. He is very con....n..ous about always doing his best work. The greatest fault of the miser is av....e. The story was told with ar....s childlike simplicity. These records are pr....ess, and must be kept with greatest care. We will t....ate these abuses no longer. It was a g...t.nous mixture, soft and elastic. The North Cape is a high p.....ry extending into the Arctic Ocean. Food in camp must be guarded against the d....d.tions of squirrels. His designs have been learned in time to f..str... them. He is a weakling, a m..ks.p afraid of all hard work. He was devoted to ph.....py, and gave much to his home town. He spoke with bitter i...y of the abuses he had seen.

This work is the crowning a.....ment of his career. The shipment is due day after t————. This article is ind.sp....ble to all householders. We cannot allow a further cont....nce of this practise. You will super.e.. him in the control of the job. Be the first to s.... these exceptional opportunities. Several men pre————ded him in this observation. A musty odor will ————nate from disused cellars. It is our b————ness to see that the work is well done. The letter was not at hand because the boy had f————ten to mail it. To su——— in life one must apply himself steadily. We should be guided by the same prin——— in all these cases. She met her fian——— at the railway station. We have received mon———s amounting to over $7000. This is a sad oc————nce, which will not happen again. The two sep——— firms will unite to form one. We are ref——— to you by your friend, Mr. John Robinson. This process will ——im.nate waste and result in great saving.

2. *Below are names of towns and cities in the United States. After each name, write the name (or abbreviation) of the State in which that town is. If the same name is given more than once, give a different state for each time it is given. Indicate only the larger of the towns so called for, not smaller ones like Salem, Maine, or Rutland, Mass.*

Skip any you do not know, and go on to the next.

Muskegon.....	Duquesne.....	Alexandria.....	Camden.....
Ansonia.....	Newton.....	South Bend.....	Rock Island.....
Walla Walla.....	Huntington.....	Paterson.....	Chattanooga.....
Hagerstown.....	Boisé.....	Jackson.....	Malden.....
Leominster.....	Washington.....	Jackson.....	Pottsville.....
Cambridge.....	Washington.....	Jackson.....	Belleville.....
Aurora.....	Erie.....	San Francisco...	Oswego.....
New Castle.....	Berkeley.....	Norwalk.....	Kenosha.....
	Topeka.....	Michigan City...	

51. For accuracy, the test-series is scored as a point scale, the sum of possible credits being 100. Each item represents a proportional part of these credits. Records were kept as

indicated for Series C in paragraphs 41 and 42. The question mark feature of the directions was without value.

52. Percentile scores for the 75 college students and 53 high school students, for the entire test-series, are as follows:

College Group				High School Group			
Point score	Per cent surpassed by same	Time of test (min.)	Per cent surpassed by same	Point score	Per cent surpassed by same	Points earned per min.	Per cent surpassed by same
41	0	26	0	31	0	1.0	0
45	3	22	9	35	9	1.5	8
50	8	20	19	40	25	2.0	28
55	23	18	43	45	45	2.5	53
60	44	16	61	50	59	3.0	73
65	65	14	76	55	79	3.5	83
70	82	12	92	60	93	4.0	91
75	94	11	96	65	98	4.5	96
82	98	7	98			5.0	96
						5.5	98

53. There is more separation of the college from the high school group in this test-series than in series C. The present test-series is of a type calculated to specially bring out the formal educational difference involved. The separation is somewhat greater in respect to total points earned than in points earned per minute.

54. The accompanying table analyses the credits earned in the items of the cued combination test, all data being combined.

Word	insure	regard	nerve	crunch	juggler	majesty	brunette	apish	sportive	hysterics	Mars	repose	shrewd	forfeit	peculiarity	coinage	mosaic
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Correct, %	91	98	98	68	68	52	85	37	18	67	91	56	80	86	69	81	68
Failure of spelling, %	3	0	0	4	21	10	11	1	1	25	2	1	11	8	6	2	1
Wrong word combined, %	5	0	1	20	8	16	1	15	36	3	2	7	3	2	1	9	20
Omitted, %	1	2	1	8	3	22	3	47	45	5	5	36	6	4	24	8	11

Word	bewail	disproportionate	dilapidated	charter	conscientious	avarice	artless	priceless	tolerate	gelatinous	promontory	depredation	frustrate	milksop	philanthropy	irony	achievement
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Correct, %	73	47	23	90	43	53	23	58	79	25	21	22	67	29	61	79	29
Failure of spelling, %	0	14	51	1	37	12	1	1	1	8	36	11	7	1	6	3	3
Wrong word combined, %	7	3	2	1	1	8	14	7	4	21	8	6	4	0	3	0	17
Omitted, %	20	36	24	9	19	27	62	34	16	46	35	61	22	70	30	18	51

Word	tomorrow	indispensable	continuance	superseded	seize	preceded	business	forgotten	succeed	principle	fiance	moneys	occurrence	separate	referred	eliminate
	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Correct, %	88	44	80	32	30	35	94	82	80	71	69	45	49	91	60	48
Failure of spelling, %	1	52	17	36	15	31	6	8	11	4	25	22	48	7	22	13
Wrong word combined, %	6	2	1	8	29	13	0	2	5	24	6	7	0	0	17	9
Omitted, %	5	2	2	24	26	21	0	8	4	1	0	26	3	2	1	30

55. Special disparity, not shown with the tables, between college and high school groups is indicated in the words *apish, gelatinous, depredation* (*s*), *achievement, seize.*

C. M. KELLEY,
F. L. WELLS.

V. GEOGRAPHICAL INFORMATION IN HIGH SCHOOL AND COLLEGE STUDENTS.

56. A list was made of towns and cities in the United States having population of over 15,000 by the 1910 census. The names were arranged in random order. The list includes 400 names. Ten alternative lists of 34 names each were prepared. One of these is the basis of the present observations, its material being reproduced with paragraph 48.

57. Responses were not scored correct unless they designated a town included in the population requirement, irrespective of there being elsewhere a smaller town as named; (cf. Salem, Maine; Rutland, Mass.; in instructions.) As a test of information, it is simply of whether the larger town in question is known; further, it is not evident whether the designations of smaller existing towns represent actual knowledge thereof, or have only chance correctness. Credit is allowable however, for such responses as Lakewood, New Jersey or Plymouth, Mass., where larger but less widely known towns have place in the list; (Lakewood, Ohio; Plymouth, Pa.).

58. The college and high school groups are considered apart, since the former is of Massachusetts, the latter of Ohio origin. This affects the results as shown in the following tables:

Town.	Muskegon	Anaonia	Walla Walla	Hagerstown	Leominster	Cambridge	Aurora	New Castle	Duquesne	Newton	Huntington	Boisé	Washington (D.C.)	Washington (Pa.)	Erie	Berkeley	Topeka	Alexandria
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Correct, % College	11	5	13	3	73	97	12	6	4	93	0	17	94	3	37	57	59	4
High School	45	0	8	0	2	81	6	17	13	0	74	21	81	8	60	9	49	0
Incorrect, % College	15	11	15	8	3	3	25	60	44	1	27	19	28		53	24	17	11
High School	19	15	24	19	7	9	91	76	21	42	13	34	30		15	30	15	32
Omitted, % College	74	83	72	89	23	0	63	34	38	5	73	64	23		9	19	24	85
High School	36	85	68	81	91	9	4	8	66	58	13	45	37		25	60	36	68

Town.	South Bend	Paterson	Jackson (Mich.)	Jackson (Miss.)	Jackson (Tenn.)	San Francisco	Norwalk	Michigan City	Camden	Rock Island	Chattanooga	Malden	Pottsville	Belleville	Oswego	Kenosha	
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	AV.
Correct, % College	29	69	4	16	11	97	13	1	40	9	53	96	11	3	33	4	32
High School	64	36	8	26	11	98	0	0	26	17	85	2	8	2	13	0	25
Incorrect, % College	21	9	..	42	..	0	20	77	41	34	32	1	20	12	24	9	24
High School	23	15	..	37	..	0	21	91	26	38	4	15	11	9	23	19	28
Omitted, % College	49	21	..	48	..	3	67	21	19	56	15	2	69	85	43	86	44
High School	13	49	..	47	..	2	79	9	47	45	11	83	81	89	64	81	47

59. Provincialism of response, favoring now the college, now the high school group, is shown in the items Muskegon, Newton, Huntington, Erie, South Bend, Rock Island, Chattanooga and Malden. This is the most important special factor in the results. Separation of college from high school as such scarcely appears outside the items Berkeley, Paterson, Oswego. The Massachusetts group is unduly favored by random selection from the large list, through the number of Boston suburbs thus included. It is indicated that the test items should not include the States where the test is made, or perhaps the States adjacent thereto. Towns listed in more than one State are also inconvenient.

60. It is suggested that the ratio of right to wrong responses has an individual significance not included in the purpose of the test, concerning the reliability of the subject's "knowledge," his critical insight into the correctness of his answers. The measure used is the per cent which the number of right responses is of the whole number of responses written. Individual differences are as shown in the accompanying distribution:

Right Answers divided by All Answers	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
College Group (75).....	1	1	2	5	4	8	12	9	6	8	5	3	6	2	2	1
High School Group (53).	2	0	7	2	7	6	8	8	4	5	4	..	..	..	..	..

61. Group difference is in the expected direction but is small in proportion to individual variability. The further education of the college group affects the criticism of this knowledge less than it does the capacity for cued combination responses.

62. Among the mechanisms of wrong responses, is first the existence of other towns having the same name as that properly designated. Jackson, N. H. (9)[6], and Camden, Maine (19) are prominent examples of this. The response Alexandria, La. (2) is virtually correct, 1915 census figures giving this town population of 18,000. (It is the only case of this kind.) Districts outside the United States are occasionally named; such responses may be fairly reasonable, as Alexandria, Egypt (2), or less so, like Topeka, China (2). Many varieties of autistic thinking are evident in the wrong responses, sound association among them. Duquesne is assisted into Iowa (7) by Des Moines and Dubuque. Boisé owes the same location (14) to the resemblance of Ia. and Ida. Jackson finds itself in Florida (59) beside Jacksonville. Norfolk brings Norwalk into Virginia (12). Rock Island proceeds via Rockland, into Maine (20), and via Block Island to New York (12); it is also found in Rhode Island (3). A peculiar feature in the results is the placement of Chattanooga in New York (16), apparently through such media as Chautauqua, Canandaigua, Cattaraugus, Canastota. Other than directly sensory associations appear responsible for the placing of Washington in Oregon (14). Berkeley, Mass. (16), has a partial determinant in the proximity to most of the subjects of Berkeley Street, Boston. Michigan City appears regularly in Michigan (112); but once out of 115 responses by 140 persons, in Indiana.

63. There is no reason to regard such material as an "intelligence" test. It applies synthetically to a restricted field of practical information. Illogical, "autistic" thought-processes are so prominent in the wrong responses as to suggest, analytically, that these reflect a capacity for critical recognition of the "real."

C. M. KELLEY,
F. L. WELLS.

[6] i. e. the response Jackson, N. H., was given nine times.

VI. AN OMNIBUS TEST FOR REASONING AND IMAGERY.

64. Performance in the present test-series, (R) is limited by reasoning and control of imagery rather than by range of information or attention to mechanical detail. The form used is reproduced herewith. The first of its tests has had con-

College Group				High School Group			
Point Score	Per cent Surpassed by same	Time of Test	Per cent Surpassed by same	Point Score	Per cent Surpassed by same	Points Earned per min.	Per cent Surpassed by same
21	0	27	0	0	0	0	0
30	0	21	1	20	6	2.5	17
40	1	19	1	30	17	5.0	43
50	9	17	7	40	27	7.5	73
60	23	15	9	50	38	10.0	91
70	24	13	22	60	53	12.5	96
80	45	11	46	70	57	15.0	98
90	66	9	74	80	72		
100	95	7	91	90	85		
		5	98	100	98		

siderable development elsewhere. The latter two are derived from the Stanford scale. It was desired to examine their adaptability to group testing. Ten alternative forms of the series were prepared.

65. In the clock tests the subject should not have access to a watch or clock face. One group of the high school students was able to see a large clock during the experiment. Their results are not improved over the others.

66. Scoring and recording were as described in paragraphs 41, 42, 51. Percentile scores for the college and high school students are as follows:

67. There is less difference of the college and high school groups than in series K, and more than in series C. As in series K, separation is more marked in total points earned than in points earned per minute.

1. *In each line every number differs from the one before it by the same amount. On each dotted line write the number that will come next in the same way.*

280, 273, 266, 259.....

378, 386, 394, 402.....

276, 280, 284, 288.....

232, 224, 216, 208.....

2. *Figure in your mind where the hands of the clock are when it is* 3:36 *o'clock. Now when the hands have changed places (so that the small hand has gone where the large hand was, and the large hand has gone where the small hand was), what time is it then?*.....

If it is	*What time is it when the hands change places*
8.13	
12.41	
10.19	

3. *A doctor must measure out exactly* 3 *ounces of medicine for a man who is sick in the woods. He has only a* 2 *ounce measure and a* 5 *ounce measure. Show how he can use these measures to get just the right dose of* 3 *ounces without any guessing. He pours from one measure to the other, and what he does not want, he pours back into his medicine bottle*.....

.....

.....

.....

.....

How would he measure out 2 *ounces with a* 3 *ounce measure and a* 4 *ounce measure.*

.....

.....

.....

How would he measure out 1 *ounce with a* 2 *ounce measure and a* 5 *ounce measure.*

.....

.....

.....

68. Accompanying table analyses the credits earned in all data obtained with this series:

Performance	Test 1	2	3
Per cent Making Perfect Score	61	50	25
½ error	..	..	22
1 "	17	32	15
1½ "	..	..	3
2 "	7	7	5
3 "	3	4	26
4 errors, and omissions	12	7	4

69. No group difficulty appears with the clock test beyond preventing the consulting of watches and clocks. In test 3, (ingenuity), failures are largely due to "guessing", i. e. such responses as "fill the 3 oz. bottle half full," and the like. It may be questioned whether more emphatic instructions against this should be given. It is believed that to be satisfied with a "guessing" response in the present experimental setting, denotes essentially significant failure to grasp the problem. The test involves more writing than is desirable, otherwise is suitable for group presentation. It shows some notable separation of college and high school groups, perfect score being made by 38 per cent of the college and 4 per cent of the high school group.

C. M. KELLEY,
F. L. WELLS.

VII. CERTAIN INDIVIDUAL RECORDS COMPARED WITH STANDARD FILE.

70. In the previous studies the chief concern has been with college and high school students, that is with persons who have not in general made an independent adjustment to life. In the present study a comparison of their records is made with a few individuals who have made such adjustments, at various levels. The findings throw light on what capacities in these test functions are compatible with different levels of vital accomplishment.

71. Attention is invited to records of the following six subjects:

		Series C		Series K	
	Subject	Total points	Points per minute	Total points	Points per minute
Per cent surpassed by...	I	6	11	10	63
" ...	II	9	28	59	85
" ...	III	18	54	97	99
" ...	IV	56	87	41	75
" ...	V	85	96	52	96
" ...	VI	93	..	100	..

72. Subjects I-III are experienced clerks working in the same office. To their employer they rank in business performance in the order represented by their test results, but these do not directly express the differences between them. This is partly because the comparative material is generally selected from the same level, of intelligence in which these subjects fall. This "dilates" the percentile differences. If the standard file had contained material from all levels of intelligence instead of from a rather high level only, these percentile scores would be higher and closer together.

Subject IV is an ambitious young man of foreign birth, an office assistant, who has some language difficulty. Subject V is a college and medical graduate; Subject VI a Phi Beta Kappa college graduate, of further academic experience.

73. These cases make relatively higher scores in points per minute than in total points earned. This betokens more hurried work. The examinations in these cases were individual, tending probably to greater haste through greater self-consciousness. It is to be supposed that the attitude of the students was influenced by the analogous situation in academic examinations and quizzes, which are less acutely hurried than psychological tests. Money prizes were offered

for the best performance in the college and high school groups, but not in the other cases.

74. Five of the more successful type of insurance salesmen, together with an office manager, were given the tests by their chief executive. The percentile scores are as follows:

		Series C		Series K		Series R	
	Subject	Total points	Points per minute	Total points	Points per minute	Total points	Points per minute
Per cent surpassed by	VII	50	12	98	41	38	32
"	VIII	5	4	46	53	5	3
"	IX	27	7	28	12	68	82
"	X	31	26	44	11	8	11
"	XI	59	55	83	52	100	96
"	XII	75	94	86	83	84	80

75. Comment of the executive making the tests, based upon the points per minute score, is in part as follows:

"If we eliminate XII, who is an office man, we find XI standing high in each of the different tests. This is in accord with what I believe to be the relation of XI's ability to the other men. In fact, I should grade them about as follows, from the standpoint of salesmanship and general usefulness to the Company in that particular: XI, VII, VIII, IX.[7]

"The difference between the five men would not be very great, although XI, perhaps, stands out somewhat more conspicuously from the group than any of the others . . .

"On the whole, it would appear that the R tests come nearer to matching up with our judgment of the men than either of the other two. Even here, however, IX takes second position, but that, after all, may be nearer right than my judgment. Five years ago, I would have placed IX at the head of the entire group, but during the last five years the character of his work has changed very considerably.

"But even in the R tests, the differences seem to me to be very much greater than actually exist. It could not be said. for example, that XI is anything like so many times as good a man as VIII [8] when it comes to the question of his value to the Company.

"The C tests work out quite well as a measure of carefulness. XII, the office man, stands at the head of the list, while the other men grade downward exactly in accord with what I believe to be the fact . . . while the R tests, if

[7] A previous rating by the same individual is XI, VII, VIII, X, IX. A rating based on 1918 new insurance produced is VII, X, XI, VIII; IX's position cannot be clearly stated, but seems much improved. This has comparatively little correspondence with the test findings, (1917).

[8] Cf. par. 72.

looked upon as a measure of resourcefulness, seem to be quite accurate with the exception of VIII, who in that quality should, I believe, rank ahead of X and perhaps ahead of VII."

76. Only partial explanation is offered of the large differences in performance between relatively equal levels of nonexperimental competence. Responses in the tests depend on what is at the surface of the conscious, or can be brought there with such effort as the subject will apply thereto. Intelligence is more closely related to the content of the conscious than to emotional or motor processes; yet there are possible many differences in the content of the conscious among persons of similar degrees of intelligence. Military psychology has shown this on a large scale, for the content of the conscious as influenced by occupation. Occupational differences can be but little invoked in the present instances. Other determinants, affecting consciousness and personality independently of intelligence, are to be sought in temperament, general interests, attitude toward the tests. These appeared especially important with Subject I who is much nearer II and III in professional usefulness than the test performance indicates.

77. Intelligence is a factor in most civilized operations, and is quicker and surer reached in formal tests than otherwise. "Other things being equal" the more intelligent will always do better mental work; and as other things are equal in chance, formal tests are at all levels better than no tests at all. They are most effective where the desideratum is a more or less routine efficiency of definite mental associations, as in clerical work. They are less effective for operations that deal with things rather than with ideas, and still less effective at the more executive levels of endeavor where less depends on intelligence in proportion to such qualities as coöperativeness, persistence, initiative, tact. "Other things being equal" they will still pick the more useful individuals; but other things are more likely to be unequal, and their inequalities more important, than at routine levels of mental work.

F. L. WELLS.

NOTES AND NEWS

SHERWIN CODY. *Commercial Tests and How to Use Them.* World Book Company, Yonkers-on-Hudson, N. Y., 1919, p. 216.

This book presents the history and technique of the National Ability Tests, which were used as the basis for the Efficiency Employment Register of high school graduates now being offered by the United States Employment Service in New York City. These Tests are intended for prospective office boys, general clerks, sales people, stenographers and bookkeepers. They are arranged in two parallel series of equal difficulty, each series containing 19 tests on such topics as: tabulation, invoicing, fundamental arithmetic, business arithmetic, writing and answering letters, stenography, copying on typewriter and on mineograph, addressing and filing envelopes, memory, English grammar, punctuation and spelling. Minute directions for administering the tests, scoring and computing the data and sample records from several school systems are given. The book should prove very helpful to those interested in educational and vocational guidance.

ARTHUR S. OTIS. *Manual of Directions for Administering the Otis Group Intelligence Scale, A Point Scale,* Edition 1919, Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y., 1919, p. 37.

This little pamphlet gives detailed directions for administering the Otis group tests, the underlying principle of which was incorporated in the United States Army tests. The complete material for giving the tests consists of (1) Examination Booklets, Form A and Form B; (2) Individual Record Card, for recording all available information about the subject tested; (3) Log Slip; (4) Examiner's Manual; and (5) Examiner's Key. The Booklets contain ten tests each of parallel difficulty, each test being suitable to subjects from the fourth grade up through the university. The tests include the following items: (1) following directions, (2) opposites, (3) disarranged sentences, (4) proverbs, (5) arithmetic, (6) geometric figures, (7) analogies, (8) similarities, (9) narrative completion, and (10) memory. Each test is arranged so that its successive elements become more and more difficult, and the score is based simply upon the total number of elements correctly answered.

WILLIAM D. LEWIS and EDGAR A. SINGER. *The Winston Simplified Dictionary,* Including all the Words in Common Use, Defined so that they can be easily Understood. (Illustrated.) The John C. Winston Company, Philadelphia and Chicago, 1919, pp. xxii+820.

The vocabulary included in this book has been chosen with special reference to educational needs and the definitions are given in plain, simple language, avoiding the frequent dictionary error of repeating in the definition the word to be defined. One or two random samples will suffice to indicate the character of the work:

Calculate (kăl'kūlāt), *v. t.* to add, subtract, multiply, or divide any sum to find the result; to determine by any process of reasoning; to estimate, as, no one can *calculate* the benefits of electricity to the world: *v. i.* to make a computation; to rely, with *or* or *upon.*

Salm-on (săm'ŭn), *n.* a sea fish, found in northern waters, which ascends rivers to lay its eggs; the yellowish-pink color of salmon flesh; *adj.* of the yellowish-pink color of salmon flesh.

The Introductory Part contains an eight-page discussion of the topic "How Our Language Has Grown," ten rules for spelling derivative words, explanatory notes on varied spelling, capitals, hyphens, etc., a short list of abbreviations used, and a guide to pronunciation. It is greatly to be hoped that this new book will supplant the cumbersome school dictionaries now in use, at least in the elementary grades, and we recommend it not only for every classroom, but for every home as well.

The General Education Board has just published a series of eight reports on various aspects of the Gary Public Schools, including the following:

ABRAHAM FLEXNER and FRANK P. BACHMAN. *The Gary Schools: A General Account*, p. 265.
GEORGE D. STRAYER and FRANK P. BACHMAN. *Organization and Administration*, p. 126, with 7 School Programs.
FRANK P. BACHMAN and RALPH BOWMAN. *Costs, School Year 1915-1916*, p. 86, with 9 Tables and 22 Schedules for Expenditures.
CHARLES R. RICHARDS, *Industrial Work*, p. 204.
EVA W. WHITE. *Household Arts*, p. 49.
LEE F. HANMER. *Physical Training and Play*, p. 35.
OTIS W. CALDWELL. *Science Teaching*, p. 125.
STUART A. COURTIS. *Measurement of Classroom Products.*

The Colorado State Teachers College Bulletin of October, 1918, consists of "A Study in Spelling," by Professor J. D. Heilman. The Bulletin of March, 1919, contains "A Study of the Mechanics of Reading," by the same author.

The Committee on Classification of Personnel in the Army has issued "A Summary of the Work of the Classification Division of the Adjutant General's Department" which contains brief descriptions of the following items: The Task of Army Personnel Organization, The Classification of Soldiers, Soldiers' Classification Card, Trade Specifications and Occupational Index, Trade Tests, Tables of Occupational Needs, Personnel Specifications, Classification and Rating of Officers, The Rating Scale, The Rating Scale Card, Relations with the Field, and The Recruit Receiving Building.

In a recent *Bulletin of the University of Wyoming* Professor June E. Downey contributes a study entitled "The Will-Profile, A Tentative Scale for Measurement of the Volitional Pattern." The profile is based upon a quantitative measurement of the following ten factors: coördination of impulses, accuracy, tenacity, resistance, assurance, motor-impulsion, speed of decision, flexibility, freedom from inertia, and speed of movement. Most of the items are determined by an examination of handwriting under various conditions; as: normal, speeded, retarded, disguised, imitated, blocked, with eyes closed, with distractions by counting by 2's or by 3's, etc. Each of the ten tests is scored on a basis of ten points, and the will-profile for each individual is constructed on the scores obtained in the ten factors.

The Bureau of Educational Experiments in New York City has secured the services of Dr. William R. P. Emerson, under whose direction a series of experiments with "nutrition classes" have been carried out. Some of the results and conclusions of this work are presented by David Mitchell in an article on "Malnutrition and Health Education" in the recent March issue of the *Pedagogical Seminary*. This paper contains a number of valuable tables on the distribution of children according to percentage of over and underweight in grades I, V, VI, VII, in a special class for exceptionally bright children, and in an open air class. After an initial measurement in height and weight these classes were subjected to several kinds of treatment; for instance, the 7th grade received instruction in health habits, physical care in the removal of defects interfering with nutrition, a half-hour rest period during the forenoon, and a mid-day lunch. The other classes received similar combinations. At the end of 19 weeks the children were remeasured to observe the changes produced by the various methods of treatment. Among the causes for lack of gain or loss in weight the following were prominent: enlarged tonsils and adenoids and carious teeth. Some of the important conclusions drawn by the author are: that provision of food in school is in itself not adequate to solve the problem of malnutrition, that stimulants, such as tea and coffee, should be eliminated, that physical defects should be removed, and that the children should be taught in hygiene and health habits.

The *Thirteenth Biennial Report* of the Board of Trustees of the Preston School of Industry, Ione, California, covering the time from July 1st, 1916 to June 30, 1918, contains an enquiry into the mental conditions, character and antecedents of 300 boys who were in the school on June 1st, 1918. The psychological data were collected by Fred. H. Allen and Warner Brown and were based upon a mental examination with the Goddard revision of the Binet scale and other supplementary tests, and upon a critical behavioristic study made during the first few weeks of the boy's stay at the school. On the basis of this study it was found that about one third of the 300 boys were dull normal or above, one third borderline cases, and the remaining third were definitely defective. Of the latter group, 27 cases showed symptoms of slight psychopathic instability and 6 cases presented conspicuous symptoms of mental abnormality.

The National Association of Employment Managers began last January the publication of a monthly bulletin, called "*Personnel.*" The second issue contains a succinct description of trade tests. Three columns in each issue are devoted to recent publications in related fields. We welcome this new publication and trust that it will contribute materially to a widespread recognition of the human factor and the psychological problems involved in the relations between employer and employee.

The University of Illinois has established a Bureau of Educational Research under the directorship of Professor B. R. Buckingham. Among the aims of this Bureau are: the compilation of bibliographies on educational topics; assisting school authorities in devising records for general and specific purposes; the study of such problems as the grouping of grades, the function of special classes, continuation and part-time schools, school statistics, and school accounting; giving aid

in selecting, retaining, supervising the teaching force; and the administration and computation of educational tests in coöperation with local school authorities.

The scope of the Bureau of Salesmanship Research at the Carnegie Institute of Technology, Pittsburgh, has recently been enlarged to include factory and clerical workers and executives, as well as salesmen. The organization is now known as the Bureau of Personnel. Its purpose is to study methods of selection, training, development, supervision, and compensation of employees.

The Public School System of Santa Ana, California, has established a Department of Research, with Mary Bess Henry as Director. The first Bulletin published by this new Department contains a discussion of "Mental Testing as an Aid in Guidance and Classification of School Children." The results reported are based on Otis' vocabulary tests, Otis' group tests of intelligence, teachers' estimates of intelligence, and scholastic records.

Professor L. L. Thurstone of the Division of Personnel and Psychology at Carnegie Institute of Technology, Pittsburgh, has recently issued two booklets for the psychological examination of college freshmen and high school seniors. Part A and Part B, and also a "Test of Engineering Aptitude for College Freshmen and High School Seniors." The former consists of one hundred short problems with directions for each. The subject is asked to solve as many as he can in twenty minutes. The engineering test contains 24 mechanical problems with directions, the subject having 30 minutes for solving as many as he can do.

Dr. Walter B. Swift, has just returned to Boston from Cleveland, where he has spent about a year installing and supervising Speech Correction in the Cleveland Public Schools. He trained up 15 teachers who are part time speech teachers. They do their regular grade work as usual. The part time speech teacher is one of the unique features of the Swift methods and systems of speech correction. In Cleveland there are now 46 classes and over 600 cases under treatment. Dr. Swift will be in Milwaukee, Wis., in July and Athens, Ohio in August. He is booked to give informal courses in these two cities this summer.

The National Society for the Study and Correction of Speech Disorder will have its summer meeting in Milwaukee, on July 4, as one of the affiliated Societies of the National Educational Association. Members of the Society and invited guests of prominence in the field of speech correction, will address the Association. Anyone interested to receive an advanced program may do so by addressing the Secretary, Miss Marguerite Franklin, 110 Bay State Rd., Boston Mass.

The following publications have been received:[1]

James Burt Miner. *Deficiency and Delinquency, An Interpretation of Mental Testing.* Educational Psychology Monographs, No. 21. Warwick and York, Baltimore, 1918, p. 355.

Roy Franklin Richardson. *The Psychology and Pedagogy of Anger.* Educational Psychology Monographs, No. 19. Warwick and York, Baltimore, 1918, p. 100.

---

[1] Mention here does not preclude further comment.

The Journal of Educational Psychology, Vol. x, No. 2, February 1919. Published by Warwick and York, Baltimore.

University of Oregon Bulletin, New Series, Vol. xv, No. 12. The University High School, issued by Department of Education. September 1918. Published by University Press, Eugene, Oregon.

Children's Year Leaflet No. 10. Advising Children in their Choice of Occupation and Supervising the Working Child. U. S. Department of Labor, Children's Bureau. Bureau Publication No. 53. Published by Government Printing Office, Washington, 1919.

Carry On, A Magazine on the Reconstruction of Disabled Soldiers and Sailors. Edited by the Office of the Surgeon General, U. S. Army. Published by the American Red Cross. Vol. 1, Nos. 7 and 8, April 1919 and May 1919.

The Porto Rico School Review. San Juan, Porto Rico. Vol. 3, Nos. 6 and 9, February and May 1919.

Revista de Psiquiatria y Disciplinas Conexas. Vol. 1, No. 3, January 1919. Published at Gremios 435, Altos-Lima, Peru.

Revista de Educacion. Tercera Epoca. Vol. 1, No. 2, March 30, 1919. Published by the Servicio Nacional de Instruccion Publica. Santo Domingo, Republica Dominicana.

Directory of American Psychological Periodicals

American Journal of Psychology—Worcester, Mass.: Florence Chandler.
Subscription $5. 600 pages annually. Edited by G. Stanley Hall.
Quarterly. General and experimental psychology. Founded 1887.

Pedagogical Seminary—Worcester, Mass.: Florence Chandler.
Subscription $5. 575 pages annually. Edited by G. Stanley Hall.
Quarterly. Pedagogy and educational psychology. Founded 1891.

Psychological Review—Princeton, N. J.: Psychological Review Company.
Subscription (with Psychological Bulletin) $6. 480 pages annually.
Bi-monthly. General. Founded 1894. Edited by Howard C. Warren.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company.
Subscription $3.25. 480 pages annually. Psychological literature.
Monthly. Founded 1904. Edited by Shepherd I. Franz.

Psychological Monographs—Princeton, N. J.: Psychological Review Company.
Subscription $4. per vol. 500 pages. Founded 1895. Edited by James R. Angell.
Published without fixed dates, each issue one or more researches.

Psychological Index—Princeton, N. J.: Psychological Review Company.
Subscription $1.25 200 pages. Founded 1895. Edited by Madison Bentley.
An annual bibliography of psychological literature.

Journal of Philosophy, Psychology and Scientific Methods—New York:
Sub-Station 84. Bi-weekly. 728 pages per volume. Founded 1904.
Subscription $3. Edited by F. J. E. Woodbridge and Wendell T. Bush.

Archives of Psychology—Sub-Station 84, New York: Archives of Psychology.
Subscription $5. 600 pp. annually. Founded 1906. Edited by R. S. Woodworth.
Published without fixed dates, each number a single experimental study.

Journal of Abnormal Psychology—Boston; Richard G. Badger. Subscription $5. 480 pages annually. Ed. by Morton Prince. Bi-monthly. Founded 1906.
Entire field of abnormal psychology.

Psychological Clinic—Philadelphia: Psychological Clinic Press.
Subscription $1.50. 280 pages annually. Edited by Lightner Witmer.
Monthly (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Training School Bulletin—Vineland, N. J.: The Training School.
Subscription $1. 160 pages annually. Ed. by E. R. Johnstone. Founded 1904.
Monthly (10 numbers). Psychology and training of defectives.

Journal of Race Development—Worcester, Mass.: Louis N. Wilson.
Subscription $3. 460 pages annually. Founded 1910. Quarterly.
Edited by George H. Blakeslee and G. Stanley Hall.

Journal of Educational Psychology—Baltimore: Warwick & York.
Subscription $3. 600 pages annually. Founded 1910.
Monthly (10 numbers). Managing Editor, J. Carleton Bell.
(Educational Psychology Monographs. Edited by Guy M. Whipple.
Published separately at varying prices. Same publishers.)

Journal of Animal Behavior—Cambridge, Mass.: Emerson Hall.
Subscription $5. 450 pages annually. Founded 1911.
Bi-monthly. Robert M. Yerkes, Managing Editor.

The Behavior Monographs—Cambridge, Mass.: Emerson Hall.
Subscription $5. 450 pages per volume. Edited by John B. Watson.
Published without fixed dates, each number a single research.

Psychoanalytic Review—Washington, D. C.: 3617 Tenth St., N. W.
Subscription $5. 500 pages annually. Psychoanalysis.
Quarterly. Founded 1913. Ed. by W. A. White and S. E. Jelliffe.

Journal of Experimental Psychology—Princeton, N. J.: Psychological Review Company. 480 pages annually. Experimental. Founded 1916.
Subscription $3.25 Bi-monthly. Edited by John B. Watson.

Journal of Applied Psychology—Worcester, Mass.: Florence Chandler.
Subscription $4. 400 pages annually. Quarterly. Founded 1917.
Edited by G. Stanley Hall. and L. R. Geissler.

THE BRANDOW PRINTING COMPANY
ALBANY, N. Y.